PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12P 19/34, 21/02, C12N 15/1
C07H 21/02

A1

(11) International Publication Number: WO 90/05785

(43) International Publication Date: 31 May 1990 (31.05.90)

(21) International Application Number: PCT/US89/05256

(22) International Filing Date: 15 November 1989 (15.11.89)

(30) Priority data:
273,455 18 November 1988 (18.11.88) US
337,601 13 April 1989 (13.04.89) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, 22nd Floor, Oakland, CA 94612 (US).

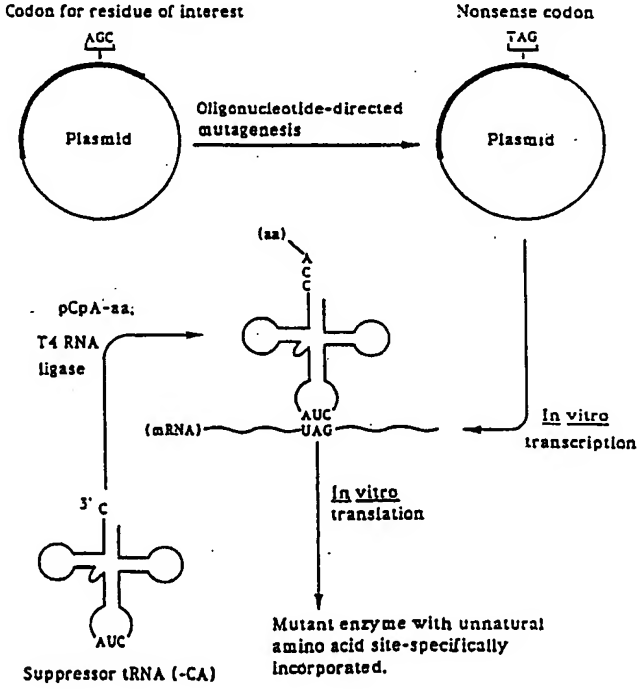
(72) Inventor: SCHULTZ, Peter ; 6331 Valley View, Oakland, CA 94611 (US).

(74) Agent: HESLIN, James, M.; Townsend and Townsend, One Market Plaza, 2000 Steuart Tower, San Francisco, CA 94105 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FI, FR (European patent), GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent), SU.

Published
With international search report.

(54) Title: METHOD FOR SITE-SPECIFICALLY INCORPORATING UNNATURAL AMINO ACIDS INTO PROTEINS



(57) Abstract

Novel methods are provided for producing proteins, containing unnatural amino acids at specific sites. The methods can utilize modified aminoacyl tRNA's capable of polymerizing a desired unnatural amino acid at unique codons within an mRNA sequence.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic	KR	Republic of Korea	SN	Senegal
CG	Congo	LI	Liechtenstein	SU	Soviet Union
CH	Switzerland	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TG	Togo
DE	Germany, Federal Republic of	MC	Monaco	US	United States of America
DK	Denmark				

METHOD FOR SITE-SPECIFICALLY INCORPORATING UNNATURAL AMINO ACIDS INTO PROTEINS

This is a continuation-in-part application of commonly assigned patent applications U.S.S.N. 273,455 and U.S.S.N. 273,786, both filed November 18, 1988, which are incorporated herein by reference.

This invention was made in part with government support under ONR Contract N00014-86-K-0522, awarded by the Office of Naval Research, and DOE grant No. DE AC03-76SF00098. The government may have certain rights in this invention.

Field of the Invention

This invention relates generally to protein biochemistry and, more particularly, to site specific modification of proteins generally useful for controlling specificity and activity of enzymes, and for altering the natural structural properties of proteins.

BACKGROUND OF THE INVENTION

Classically, biochemists have modified protein function by either chemically altering isolated proteins or by selecting naturally occurring variants. Chemical modification is typically directed to unusually reactive and solvent accessible amino acid side chains, but often the desired modification sites are not accessible or the modifications of interest are not chemically feasible. Low specificity of modification reactions dramatically hinders the usefulness of this approach. Moreover, naturally occurring variants are rare, usually require substantial analysis to determine the nature of any variation, and are generally limited to substitutions with naturally occurring amino acid residues.

With the advent of molecular biology technology, other approaches have been developed which allow, albeit in a limited manner, particular amino acids to be incorporated

into proteins or peptides during the process of peptide bond polymerization. Peptide synthesis and semisynthetic methods have been used to introduce novel amino acids into very small proteins and peptides. Modified amino acids have been uniformly incorporated into peptides and proteins using functional analogues of aminoacyl transfer RNA's (tRNA's). In addition, several unnatural amino acids have been incorporated into dipeptides using chemically misacylated tRNA's.

All of these methods suffer from one or more of the following drawbacks: lack of site specificity in the introduction of the novel amino acid; heterogeneity in sites of modification; low efficiency in modifications; a requirement for extensive characterization to determine precisely the location and nature of any substitution; severe size restrictions on the protein of interest and very limited possibilities for substitutions or modifications. Thus, there exists a need for improved methods for generating proteins with specific modifications at desired locations. The present invention fulfills these and other needs.

SUMMARY OF THE INVENTION

In accordance with the present invention, novel methods are provided for site specifically incorporating an unnatural amino acid analogue into a protein, the methods comprising the steps of:

(a) introducing a preselected codon into at least one site in a mRNA sequence encoding the protein; and

(b) translating the mRNA sequence in a protein synthesizing system comprising an aminoacyl tRNA analogue capable of polymerizing the unnatural amino acid analogue into a nascent polypeptide chain on direction of the preselected codon.

The protein synthesizing system is preferably an <u>in vitro</u> protein synthesizing system, and the preselected codon a

termination codon, such as UAG (amber), inserted at predetermined sites.

The unnatural amino acid analogue will typically be selected from modified natural amino acids, modified uncharged amino acids, modified acidic amino acids, modified basic amino acids, non-alpha amino acids, amino acids with altered ψ, ϕ angles, and amino acids containing functional groups selected from the group of nitro, amidine, hydroxylamine, quinone, aliphatic, cyclic and unsaturated chemical groups. Preferably, the aminoacyl tRNA analogue is the only aminoacyl tRNA molecule in the protein synthesizing system capable of recognizing the preselected codon and the preselected codon is introduced into one site of the mRNA sequence encoding the protein which typically has a molecular weight greater than about ten thousand daltons. The unnatural amino acid analogue may be situated within about 100 angstroms of a substrate binding site, an enzymatic active site, a protein-protein interface, a cofactor binding site, or a ligand (agonist or antagonist) binding site.

Another aspect of the present invention includes novel proteins (usually greater than about 10 Kd) that are stoichiometrically substituted at one or more predetermined sites, preferably substantially homogeneously, with an unnatural amino acid analogue. Analyzing the physical or biochemical properties of the protein can determine various properties, such as static physical properties of the polypeptide chain, mechanism of action of an enzymatic reaction, specificity of protein binding to ligand, dynamic interaction of amino acid residues of a subject protein with a substrate, folding of the protein, or interaction of the proein with other proteins, with nucleic acids or with sugars. Typically, the protein is analyzed within about 100 angstroms of the unnatural amino acid analogue insertion.

In another aspect, the present invention provides methods for making multiple alternative substitutions at preselected amino acid positions of a protein comprising the steps of:

a) producing one mRNA with mistranslation codons at sites in the mRNA corresponding to the preselected amino acid positions; and

b) translating the mRNA in a series of two or more translation systems each comprising an aminoacyl tRNA analogue, whereby the protein produced by one translation system differs from the protein produced by another system at the preselected amino acid position.

Preferably, one amino acid position is substituted and the difference between proteins produced by the different translation systems is predetermined by the preselection of unnatural amino acid analogues attached to the aminoacyl tRNAs. The unnatural amino acid substitution may be, for example, D-phenylalanine, (S)-p-nitrophenylalanine, (S)-homophenylalanine, (S)-p-fluorophenylalanine, (S)-3-amino-2-benzylpropionic acid, or (S)-2-hydroxy-3-phenylpropionic acid.

Yet another aspect of the present invention relates to methods for producing an aminoacyl tRNA analogue molecule comprising the steps of:

a) attaching a predetermined unnatural amino acid analogue by an aminoacyl linkage at 2' or 3' ribosyl hydroxyl positions on the 3' terminal nucleotide of a multi-nucleotide molecule (MNM); and

b) ligating the aminoacyl-multi-nucleotide molecule (aminoacyl-MNM) to a truncated tRNA molecule (tRNA(-Z)), wherein a functional aminoacyl tRNA analogue molecule is formed.

Preferably, nucleotide molecule (MNM), which may be a dinucleotide such as 5'-pCpA-3', corresponds to a tRNA 3' terminus. The ligation of the multi-nucleotide molecule (MNM) to the tRNA(-Z) molecule typically generates a complete tRNA molecule, and the tRNA(-Z) may be derived from a run-off transcript.

The attaching of the predetermined unnatural amino acid analogue by an aminoacyl linkage at 2 or 3' ribosyl

hydroxyl positions on the 3' terminal nucleotide of a multi-nucleotide molecule (MNM) is preferably accomplished by the steps of:

a) protecting reactive chemical groups of the MNM with protective agents;

b) protecting reactive non-aminoacyl reactive groups of the amino acid analogue with a blocking agent;

c) acylating the MNM with a blocking agent-protected amino acid analogue; and

d) removing the protective agents and blocking agents from the protected reactive sites.

Some or all of the reactive group protecting steps are substituted with steps using blocking or protective agents selected from the group consisting of: o-nitrophenylsulfenyl (NSP); β-cyanoethyl (EtCNO); benzyloxycarbonyl (CBZ); 9-fluorenylamethyloxycarbonyl (FMOC); 2-(4-biphenyl) isopropyloxycarbonyl (BPOC); vinyloxycarbonyl (VOC); tetrahydropyranyl (THP); methoxytetrahydropyranyl; and photolabile groups, including 4-methoxy-2-nitrobenzyloxycarbamates (NVOC). The protecting steps may be performed using o-nitrophenylsulfenyl (NPS) for both the blocking agents and protective agents, and the ligating of the aminoacyl-MNM to the tRNA(-Z) may be performed by the enzyme T4 RNA ligase.

Another aspect of the present invention comprises aminoacyl tRNA analogues having the formula X - A - Y - M, wherein:

X = 5' nucleotide sequence of a tRNA molecule;

A = anticodon nucleotides;

Y = 3' nucleotide sequence of a tRNA molecule, such as 5'-pCpCpA-3';

M = amino acid analogue selected from the group consisting of:

i) modified uncharged natural amino acids;

ii) modified acidic natural amino acids; and

iii) non-alpha amino acids.

These analogues will be able to direct the polymerization of the M component into a nascent polypeptide chain and can serve as an acceptor for further peptide polymerization. For example, an analogue corresponding to $tRNA^{Phe}_{CUA}$ aminoacylated with (S)-p-nitrophenylalanine can be produced, wherein:

a) X comprises the 5' segment of $tRNA^{Phe}_{CUA}$ containing a "D loop" and part of an "anticodon loop";

b) A (anticodon) comprises the trinucleotide 5'-pCpUpA-3';

c) Y comprises the 3' segment of $tRNA^{Phe}_{CUA}$ containing part of an "anticodon loop", a "variable loop", a "TΨC loop", and an "acceptor stem"; and

d) M is (S)-p-nitrophenylalanine.

The present invention further includes translation systems comprising such aminoacyl tRNA analogues. Also included is a coupled transcription and translation system, wherein products of the transcription system are translated by a translation system comprising such aminoacyl tRNA analogues.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of the method for introducing unnatural amino acids site specifically into proteins.

Figure 2 shows schemes for synthesizing aminoacylated pCpA.

Figure 3 shows the construction of the plasmid pSG7, the vector for in vitro expression of β-lactamase. Segment a is the 259-bp *Bam*HI-*Eco*RI fragment of pKKK223-3 (Brosius and Holy, (1984) Proc. Natl. Acad. Sci. USA, 81:6929), containing the tac promoter. Segment b is the 376-bp *Ssp*I (linkered to *Eco*RI-*Pvu*I fragment of pTG2*del*1

(Kadonaga et al., (1984) J. Biol. Chem. 259:2149) containing the first part of the gene for RTEM β-lactamase (Sutcliffe, (1978) Proc. Natl. Acad. Sci. USA, 75:3737; Ambler and Scott, (1978) Proc. Natl. Acad. Sci. USA, 75:3732; and Pollitt and Zalkin, (1983) J. Bacteriol. 153:27) with a 63-bp deletion corresponding to 21 amino acids in the leader sequence. Segment c is the 1386-bp *Pvu*I-*Hae*II fragment of pT7-3 (Tabor and Richardson, (1985) Proc. Natl. Acad. Sci. USA, 82:1074), containing the remainder of the β-lactamase gene and the ColE1 origin of replication. Segment d is the 1430-bp *Hae*II-*Hae*I fragment of *pGP1-2* (Tabor and Richardson, (1985) Proc. Natl. Acad. Sci. USA, 82:1074), containing the kanamycin resistance gene from Tn*903* (Oka et al., (1981) J. Mol. Biol. 147:217). This gene is oriented so as not to be under the transcriptional control of the tac promoter. Segment e is the 289-bp *Hae*II-*Pvu*II (ligated to the blunt-ended *Bam*HI site of segment a to regenerate only the *Bam*HI site) fragment from pT7-3.

Mutants at Phe66 (*on figure) were generated using the method of Eckstein (Nakamaye and Eckstein, (1986) Nucl. Acids. Res., 14:9679). The 204-bp *Eco*RI-*Hinc*II fragment, containing the codon for Phe66, was cloned into M13mp18. Three synthetic oligodeoxymucleotides, 5'-ATCATTGGATAACGTTCTT-3', 5'-ATCATTGGAGCACGTTCTT-3', and 5'-ATCATTGGCTAACGTTCTT-3' (underlined bases denote mismatches to the wild-type sequence) were used to generate the F66Y, F66A, and F66am mutants, respectively. Mutagenesis efficiencies were 100% for Tyr66, 83% for Ala66, and 60% for TAG66.

Figure 4 shows the in vitro synthesis and purification of truncated β-lactamase. *Lane 1:* Crude in vitro reaction; *Lane 2:* Purified β-lacatamase synthesized in vitro; *Lane 3:* Purified β-lactamase synthesized in vivo (JM101/pSG7).

Figure 5 shows the tRNA PHE/CUA(-CA) sequence as determined by the enzymatic method. The $tRNA^{Phe}_{CUA}$(-CA) was 3' end-labelled with [5'-^{32}P]-pCp and the sequence was determined by the enzymatic method (Donis-Keller, (1980) Nucl. Acids. Res., 8:3133). The products of the digestion

reactions were loaded onto a 10% denaturing polyacrylamide gel. An autoradiogram of the sequencing gel is shown: no enzyme (lane 1), -OH digest (lanes 2 and 7), RNase T1 (G-specific, lane 3), RNase U2 (A-specific, lane 4), RNase Phy M (U+A-specific, lane 5) RNase B. cereus (U+C-specific, lane 6). The sequence of the anticodon stem and loop is shown, the site of CUAA incorporation is indicated by larger letters. $tRNA^{Phe}_{CUA}$(-CA) was either purified by preparative gel electrophoresis and used in chemical misacylation reactions, or treated with nucleotidyl transferase (Cudney and Deutscher, (1986) J. Biol. Chem., 261:6450), gel-purified and used in misacylation reactions with yeast PRS.

Figure 6 shows the test of acylated and nonacylated suppressor tRNA in vitro. Reactions (30μL) were carried out as described in Figure 4, cooled to 0°C and centrifuged. Three μL of each supernatant was denatured and loaded onto a 12.5% SDS polyacrylamide gel (Laemmli, (1970) Nature (London), 227:680), which was dried and autoradiographed following electrophoresis. *Lane 1:* Reaction primed with pSG7 (truncated β-lactamase); *Lane 2:* Reaction primed with pF66am, with no added suppressor; *Lane 3:* Reaction primed with pF66am, with non-acylated suppressor (5μg) added. Lanes 1, 2 and 3 were supplemented with [^{3}H]-Phe (Amersham) to a final specific activity of 190 Ci/mol; *Lane 4:* Reaction primed with pF66am and 5μg suppressor that had been acylated enzymatically with [^{3}H]-Phe (specific activity 9.4 Ci/mmol Phe-tRNA).

Enzymatic misacylation reactions (300 μL total volume) contained the following: 4 μM $tRNA^{Phe}_{CUA}$ (30 μg, which had been desalted and lyophilized following gel purification), 80 μM phenylalanine, 40 mM Tris-HCl (pH 8.5), 15 mM $MgCl_2$, 45 μg/mL BSA, 3.3 mM DTT, 2 mM ATP and 22 Units yeast PRS (where 1 unit activity incorporates 100 pmol Phe in 2 minutes at 37°C under the following conditions: 2 μM $tRNA^{Phe}$(Boehringer Mannheim), 2 mM ATP, 3.3 mM DDT, 8.1 μM Phe, 40 mM Na HEPES (pH 7.4), 15 mM $MgCl_2$, 25 mM KCl, and 50 μg/mL BSA). The reaction mixture was incubated at 37°C for 3 minutes, then quenched by addition of 2.5 M NaOAc (pH 4.5) to

10% v/v. The quenched reaction was immediately extracted with phenol (pre-equilibrated with 0.25 M NaOAc, pH 4.5), phenol:$CHCl_3$(1:1), $CHCl_3$, then precipitated with EtOH. The extraction and precipitation were repeated twice more. The tRNA was then desalted on a Pharmacia fast desalting column and lyophilized. The lyophilized mixture of acylated and non-acylated tRNA was stored at -80°C until immediately prior to its use in in vitro protein synthesis reactions.

β-Lactamase activity in the supernatants of these reactions was measured using the nitrocefin hydrolysis assay (O'Callaghan et al., (1972) Antimicrob. Ag. Chemother., 1:283. One nitrocefin hydrolysis unit (1 μmole nitrocefin hydrolyzed/min/mL, 0.1 mM nitrocefin, 50 mM phosphate buffer, pH 7) corresponds to 0.61 μg enzyme, as determined by Bradford assay.).

Reaction		μg/mL
1	pSG7	44.6
2	pF66am	0
3	pF66am, non-acylated suppressor	0
4	pF66am, acylated suppressor	6.7

Figure 7 shows the method of chemical aminoacylation of the dinucleotide pCpA. The dinucleotide pCpA was prepared by standard solution phase phosphotriester synthesis (Jones et al., (1980) Tetrahedron, 36:3015; Van Boom and Wreesman in "Oligonucleotide Synthesis", Gait (Ed.), IRL Press, Washington, 1984). The fully protected molecule was 4-chlorophenyl-4-N-anisoyl-2'-O-tetrahydropyranyl-5'-O-[β-cyanoethyloxyphosphoryl] cytidylyl (3'-5')-[6-N, 6-N, 2'-O, 3'-O-tetrabenzoly] adenosine. Then o-Nitrophenylsulfenyl chloride (1.8 mmol) and triethylamine (1.8 mmol) were added over six hours to pCpA (285 μmol) dissolved in dimethylsulfoxide (68 mL). The reaction was quenched by addition of 50mM ammonium acetate, pH 5 (100 mL) and the solvent was removed in vacuo. Purification by reverse phase

HPLC (HPLC conditions: A = 5 mM ammonium acetate pH 5, B = MeCN, gradient = 0 to 15% B in 60 min.; 15 to 30% B in 30 min. flow rate = 8 mL/min., column = Whatman Partisil 10 M-20 10/50 ODS-3.) yield NPS-pCpA in 65% yield and recovered pCpA in 22% yield. NPS-pCpA was desalted by reverse phase HPLC (HPLC conditions: A = H_2O, B = MeCN, gradient = 0 to 15% B in 60 min., 15 to 30% B in 30 min., flow rate = 8 mL/min., column = Whatman Partisil 10 M-20 10/50 ODS-3.) followed by passage through a Dowex column (Li+ form). o-Nitrophenylsulfenylphenylalanine (74 μmol) and N,N'-carbonyldiimidazole (83μmol) were stirred under nitrogen in anhydrous dimethylsulfoxide (320 μL) for thirty minutes. The solution was then added to the lithium salt of NPS-pCpA (15 μmol, dried by repeated co-evaporation with toluene). The reaction was stirred under nitrogen at 50°C for eight hours, then quenched at 0°C by addition of 50 mM ammonium acetate, pH 5 (2 mL). Lyophilization followed by reverse phase HPLC (HPLC conditions: A = 50 mM ammonium acetate, B = MeCN, gradient = 0 to 70% B in 70 min., flow rate = 8 mL/min., column = Whatman Partisil 10 M-20 10/50 ODS-3.) provided the desired product in 16% yield with 38% starting material being recovered. After lyophilization, the product (2.4 μmol) was dissolved in 40 mM sodium thiosulfate, 50 mM sodium acetate, pH 4.5 (2 mL) and stirred for one hour. Reverse phase HPLC (HPLC conditions: A = 8 mM HOAc, B = MeCN, gradient = 0 to 30% B in 45 min., flow rate = 4 mL/min., column = Whatman Partisil 10 M9/50 ODS-3.) afforded the deprotected acyl pCpA in 81% yield. All products were characterized by UV, NMR and 2D NMR.

Figure 8 shows purification of Phe66 β-lactamase synthesized according to Figure 1. β-lactamase was purified from 900 μL pF66am-primed reaction that had been supplemented with 150 μg chemically acylated Phe-$tRNA_{CUA}$ following the procedure described in Figure 4. Typical yields were 0.3-0.7 μg (7-15%) of purified enzyme, starting from 4.5 μg in crude reaction.

Samples (50-200 ng/band) were electrophoresed on a 12.5% SDS-polyacrylamide gel (Laemmli, (1970) Nature

(London), 227:680), which was subsequently silver stained. The diffuse bands at M_r=67000 and 60000 are artifacts commonly observed during high sensitivity silver staining (Merril et al., (1983) Methods Enzymol., 96:230). *Lane 1:* Crude in vitro reaction; *Lane 2:* Purified in vitro β-lactamase; *Lane 3:* Purified in vivo β-lactamase (JM101/pSG7).

Figure 9 shows tryptic peptide mapping of tryptic digest and peptide mapping of wild-type and suppressed β-lactamase. Wild-type β-lactamase was uniformly labelled with [^{3}H]-phenylalanine by in vitro protein synthesis from pSG7 in the presence of added [^{3}H]-phenylalanine. Non-labelled β-lactamase was added to the products of the in vitro synthesis prior to purification of the enzyme by gel filtration on sephadex G-75 (Pharmacia) and chromatofocusing chromatography as described in Figure 4. The sequence of β-lactamase from pSG7 is shown, trypsin cleavage sites are indicated by spaces, peptides containing Phe are indicated by bold type, Phe 66 is underlined: MSHPETLVK VK DAEDQLGAR VGYIELDLNSGK ILESFRPEER FPMMSTFK VLLCGAVLSR VDAGQEQLGR R IHYSQNDLVEYSPVTEK HLTDGMTVR ELCSAAITMSDNTAANLLLTTIGGPK ELTAFLHNMGDHVTR LDR WEPELNEAIPNDER DTTMPAAMATTLR K LLTGELLTLASR QQLIDWMEADK VAGPLLR SALPAGWFIADK SGAGER GSR GHAALGPDGKPSR IVVIYTTGSQATMDER NR QIAEIGASLIK HW. The tryptic peptides were separated using a Pharmacia Pep RPC 5/5 column. The absorbance of the column effluent was monitored at 254 nm (panel A) and fractions of 0.5 mL were collected and counted (panel B). Radioactive suppressed β-lactamase was synthesized in vitro from pF66am in the presence of added [^{3}H]-Phe, $tRNA^{Phe}_{CUA}$. The purification and trypsin digestion of the suppressed β-lactamase were identical except that only 6,000 CPM of labelled suppressed enzyme were used in digestion reactions (panel C).

Figure 10 shows the characterization of native and mutant β-lactamase. Wild-type and Phe 66-suppressed β-lactamase were purified to homogeneity from 1mL in vitro reactions primed with pSG7 and pF66am/Phe-$tRNA_{CUA}$, respectively. Initial rates of nitrocefin hydrolysis were determined, at 24°C in 50 mM sodium phosphate, pH 7/0.5%

DMSO, at substrate concentrations ranging from 25-250 μM. K_M and V_{max} values were obtained from Eadie-Hofstee plots, and Bradford assay quantitations of the enzymes were used to determine k_{cat} values.

Kinetic parameters for the mutant enzymes were determined as follows: In vitro reactions (60μl) containing 15 μCi[^{35}S]-methionine (Amersham) were primed with pF66Y, pF66am/Phe-tRNA$^{Phe}_{CUA}$, pF66am/p-FPhe-tRNA$^{Phe}_{CUA}$, pF66am/p-NO_2-Phe-tRNA$^{Phe}_{CUA}$, pF66am/HPhe-tRNA$^{Phe}_{CUA}$, pF66am/PLA-tRNA$^{Phe}_{CUA}$, pF66am/ABPA-tRNA$^{Phe}_{CUA}$, or pF66am/D-Phe-tRNA$^{Phe}_{CUA}$ and incubated at 37°C for 30 min. K_M and V_{max} values were determined as described above using crude enzyme immediately following incubation. Quantitation of the enzymes was achieved by first precipitation the crude reaction with trichloroacetic acid and washing at 90°C to remove unincorporated label (Pratt in "Transcription and Translation", Hanes and Higgins (Eds.), IRL Press, Oxford, 1984, pp.179-209). The measured incorporated radioactivity for the Phe66 enzyme was then used, together with the k_{cat} = 880s^{-1} determined from Bradford assay quantitation, to calculate the amount of incorporated radioactivity/μmol of enzyme (typically 5 mCi/μmol). This value was then used to quantitate the mutant enzymes, based on the incorporated radioactivity measured for each. Values shown are the averages of three determinations.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides novel methods for synthesizing proteins containing unnatural amino acids at specific sites. The methods preferably utilize modified aminoacyl tRNA's capable of polymerizing the desired unnatural amino acid(s) at unique codon(s) within an mRNA sequence. Utilizing these methods, a wide variety of unnatural amino acids may be selectively introduced into proteins of interest. The methods can provide proteins which are substantially homogeneously substituted at selected sites in stoichiometric amounts. The procedures are inherently simple and allow control of both the type and site of

modifications to a protein molecule within a virtually limitless number of possible variations.

One aspect of the invention relates to the production of modified tRNA molecules and their use in producing desired proteins as follows:

a) preparing a nucleic acid sequence capable of being translated into a desired polypeptide, the nucleic acid sequence including at least one codon which will be dedicated to a desired preselected amino acid substitution within the polypeptide;

b) obtaining or synthesizing an aminoacyl tRNA analogue which will recognize the dedicated codon and function as an adaptor molecule to direct the polymerization of the amino acid substitution into the polypeptide;

c) combining the nucleic acid sequence with a protein translation system containing the aminoacyl tRNA analogue, whereby the translation system will function to normally translate the nucleic acid message, except that the aminoacyl tRNA analogue will direct the incorporation of the amino acid substitution for the otherwise naturally occurring corresponding natural amino acid; and

d) allowing the translation system to function so the sequence will be translated and the system will substitute at the direction of the selected codon the corresponding predetermined amino acid analogue into the resultant protein.

Although each of these steps relate to important parts of the invention, various modifications will be readily apparent to one skilled in the art to adapt the procedures to particular specific uses as detailed below.

Proteins are fundamental building blocks of living organisms and serve multiple functions. Typically they serve structural functions, catalytic (or enzymatic) functions or a mixture of the two. Proteins are synthesized on ribosomes which polymerize polypeptide chains out of a set of 20 common amino acid monomers according to the information contained in the sequence of nucleotides making up the "messenger RNA" (mRNA). The mRNA is "translated" by the ribosomes which

"read" three-nucleotide segments (one codon) at a time. From a particular AUG, or initiation codon, the ribosomes read successively in three-nucleotide segments, establishing the "frame" of translation.

RNA is composed of 4 different types of nucleotides containing the adenine, cytosine, guanine and uracil. In three-nucleotide segments (codons), there are 64 possible sequence combinations, three of which normally will not translate and result in termination of further polypeptide elongation. These three codons, UAG, UAA and UGA, are the normal termination codons. The other 61 codons have corresponding adaptor molecules (tRNA's) which recognize (or match) the message codon by complementarily matching with these bases. The complementary sequence is contained in this "anticodon" of the tRNA. Another segment of this tRNA adapter molecule serves to position an amino acid at the correct site in the ribosome to serve as a substrate for the "elongation" reaction, whereby the nascent chain is polymerized to the aminoacyl moiety on the tRNA. The 5' terminal codon codes for the amino terminal amino acid, and successive codons direct the successive carboxy addition of the next amino acid in the nascent chain. Thus, the polypeptide chain is synthesized beginning at the amino terminus, with each subsequent amino acid added at the carboxy terminus.

Normally, the tRNA is enzymatically "charged" with the correct amino acid moiety with extremely high fidelity so that the adaptor molecule has the correct amino acid attached which properly corresponds to the anticodon. If a tRNA is "mischarged", that tRNA will properly recognize the anticodon and the properly positioned; but improperly acylated amino acid moiety will, nevertheless, be polymerized into the nascent polypeptide chain. Furthermore, this can be extended to a tRNA which has a modified anticodon matching a termination codon. These are known as "suppressor" tRNA's, because they suppress the effect of in-frame chain termination codons which may have been introduced into a

message. This phenomenon is, in part, a fundamental basis of this invention.

This invention uses processes and molecules which, in many cases, have not been uniquely defined chemically, and uses general terms which do not necessarily match the uses precisely the same by some in the field. The following definitions are primarily based on functionalities. Much of the state of the art and concepts utilized here are contained in Watson et al., (1987) Molecular Biology of the Gene, Vols. 1 and 2, hereafter referred to as Watson et al., Gene, specifically herein incorporated by reference.

The term "reading" is the process by which the translation system recognizes a given codon of the message and polymerizes, at the direction of that codon, a particular amino acid into the corresponding position of the nascent polypeptide chain.

The term "misreading" is used to refer to mistranslation relative to the code of correspondence between the codon and the amino acid inserted into the nascent polypeptide chain synthesized by the original translation system (i.e., before the selected aminoacyl tRNA is otherwise incorporated into the translation system).

The term "termination codon" refers to the codons normally used to signal translation termination in the translation system of use. Where a natural source for the translation system is used, these will typically be the codons utilized in the "universal code". Normally these are the codons UGA, UAG and UAA. However, it is possible to generate translation systems with an entirely different correspondence of codon with amino acid, and so the term is also extended to include whatever codon is used in the system being utilized.

The term "tRNA analogue" refers to any molecule which is an analogue of a tRNA with respect to the activities of nascent peptide chain translocation and codon recognition. Each particular tRNA species is not a definitive homogeneous chemical entity, since numerous methylation or other modifications or changes in the primary nucleic acid sequence

may be made which may have minor or no effect on its essential properties. The term is here broadened beyond its use to indicate a nebulously defined core chemical entity, including various modified forms thereof. Generally, all tRNA molecules which function to recognize a specific codon and are transcribed directly from a single gene will be considered collectively as "a tRNA". Herein, the functional definition is more relevant than a chemical description.

While many tRNAs from various sources have been defined in a general sense according to the "core" nucleotide backbone sequence (Sprinzel et al., (1987) Nucleic Acid Research 15:R53; GenBank™/EMBL DataBank), the number and sites of methylations and other modifications may be heterogeneous or imprecisely defined. This definition is specifically intended to include each variant of a heterogeneous or homogeneous category of molecules containing minor modifications of a known tRNA including, but not limited to, differences in the methylation or other modification patterns, differences in the nucleic acid sequence of the tRNA backbone (including substitutions, additions, deletions, and modified bases), tRNAs from exogenous sources, and other molecules which may have relevant functions common to tRNA molecules. The efficiency of the interactions between the translational components (e.g., ribosome, elongation factors, tRNA's) need not be especially high, but a person of ordinary skill in the art will be able to recognize the essential functions relevant to the invention in each of its various embodiments. These minimal tRNA functionalities require that the molecule may be acylated by some process, enzymatic or chemical, and that the acylated molecule have adapter molecule activity. The kinetics of interactions are not especially critical but may be important in terms of efficiency.

The term "aminoacyl tRNA analogue" refers to any analogue of an aminoacyl tRNA molecule which:

a) functions as an adapter molecule, in such a manner that it will appropriately interact with a messenger

RNA, a ribosome, associated translation and elongation factors and the nascent polypeptide chain; and

b) which contains:

i) a functional anticodon entity and

ii) an amino acid analogue moiety that can be polymerized into the nascent polypeptide chain.

It should also be noted that in some procedures based on the present invention, it may be sufficient that the amino acid analogue be the terminal amino acid and, thus, polymerization into the nascent polypeptide molecule need not necessarily imply that the amino acid moiety itself be capable of accepting further amino acid monomers (i.e., further peptide elongation).

Chemically, an aminoacyl tRNA may be defined as a molecule comprising:

X - A - Y - B - M;

where X is the 5' segment of a tRNA, consisting of nucleotides and modified nucleotides (methylations and other modifications on the base components) making up the "D loop" and part (to the anticodon entity) of the "anticodon loop";

A is the anticodon segment of the tRNA; narrowly defined as the 3 nucleotides which match with the codon to be translated, but may be extended to include adjacent nucleotides within 3 nucleotides of the anticodon;

Y is the 3' segment of a tRNA, consisting of nucleotides and modified nucleotides (methylations and modifications on the base components) making up part of the "anticodon loop" and the "variable" and "TΨC" loops and acceptor stem;

B is the 5'-pCpCpA-3' terminus of the tRNA, normally not coded by the tRNA gene and added on by the 3' tRNA nucleotidyltransferase; and

M is the amino acid moiety.

Note that this definition is not meant to exclude the possibility of aminoacylating a shortened tRNA molecule with 3' terminal nucleotides removed, such as tRNA(-A), tRNA

(-CA) or tRNA(-CCA). If functional, they would be equivalent to a tRNA as used herein.

The term "multi-nucleotide" (MNM) refers to a short segment of nucleic acid, typically ribonucleic acid. The term is used in the context of the preparation of an aminoacyl tRNA analogue. The corresponding tRNA (i.e., the deacylated form) is shortened by removal of a short segment (Z) to form the "tRNA(-Z)" or tRNA molecule with the corresponding Z segment removed. Alternatively, a truncated tRNA(-Z) can be generated directly by recombinant DNA or chemical methods. The Z segment corresponds in some sense, to the multi-nucleotide(MNM). Thus, upon ligation of the shortened tRNA(-Z) to the multi-nucleotide molecule (MNM), a tRNA results which is equivalent to the deacylated aminoacyl tRNA analogue. One method of the invention uses aminoacylated-MNM's as substrates for ligation of tRNA(-Z) molecules to form aminoacyl tRNA analogues.

Thus, the term "tRNA(-Z)" refers to that molecule which is ligated to the aminoacyl-multi-nucleotide to produce a functional aminoacyl tRNA analogue molecule. It is particularly intended to include the component which is a shortened form of a tRNA, typically with a few of the 3' terminal nucleotides removed. The ligation of the tRNA(-Z) to the aminoacyl-multi-nucleotide (aminoacyl-MNM) generates a molecule which will become (or is) a functional aminoacyl tRNA analogue. Typically Z and MNM are equivalent and, in the preferred embodiment, will be the 5'-pCpA-3' dinucleotide.

The term "unnatural amino acid analogue" refers to a molecule that is either directly an analogue or modification of an amino acid. It would include modified natural amino acids, unnatural amino acids, analogues of amino acids and derivatives of amino acids.

The set of natural amino acids would include those amino acids that are commonly used in the polymerization process performed by ribosomes. Normally, these amino acids have codons which operate to signal for polymerization into protein. Although unusual amino acids exist naturally in

proteins, they are usually relatively simple modifications of members of the group of twenty common amino acids. Also included would be amino acids which actually do occur in nature, but are not polymerized in their final form during the polymerization (or translation) process. These natural modifications apparently result from post-polymerization modification of the amino acid that is performed either in the nascent chain stage, or more probably, upon completion of the polypeptide chain. These include the post-translationally modified amino acids such as 4-hydroxyproline, 5-hydroxylysine, cystine and others. These modifications are made enzymatically and are highly restricted both in the type of modification made and the amino acid which is targeted for modification.

Modified natural amino acids, unnatural amino acids, analogues of amino acids and derivatives of amino acids are intended to include all functional modifications of analogues of amino acids, both alpha and otherwise. This would also include modifications which involve substitution or addition of unusual atoms, addition of side groups including cofactors or their binding sites, glycosylations and acetylations.

The term "preselected codon" refers to a codon which is intended to be changed and will, in some functional form, be within the reading frame of the protein to be produced. Thus, if a particular sequence has more than one reading frame, the codon need only be a change intended to affect one of them.

The term "site specific incorporation" refers to the introduction into known sites of either a particular codon into a specific site in the reading frame of a message, or of a particular amino acid analogue into a specific site in a polypeptide chain. It will be recognized that since there is a one to one positional correspondence between codon positions and their integrated amino acid sites, the site of an amino acid analogue substitution is determined by its corresponding codon position. Consequently, site specificity

of amino acids may be derived from determination of the codon site, and vice versa.

The nascent polypeptide chain is the incompleted polypeptide chain resulting from the translation of the mRNA which is 5' proximate to the current codon. The current codon "directs" the specificity of the next amino acid analogue which is to be polymerized in the nascent chain. In the normal elongation process, the nascent polypeptide chain is polymerized onto the aminoacyl moiety attached to the tRNA which recognizes the codon adjacent to the A site of the ribosome.

Use of the terms protein and polypeptide is intended to include the products of the system that are modified molecules substantially equivalent to a protein or polypeptide. This is meant specifically to include a protein or polypeptide, as well as both its apoprotein and holoprotein forms. Included in the definition are polypeptide molecules:

a) having a modified amino acid substituted at the normal site of an amino acid (equivalent to a modified amino acid); or

b) having an amino acid like moiety which may differ in structure or composition, including, but not limited to:

i) a moiety which would have the peptidyl linkage involving an amino group off a beta, gamma, delta or other carbon atom (i.e., non-alpha amino acid);

ii) a moiety which might have an atom other than a carbon or nitrogen atom along the polypeptide backbone;

iii) an amino acid containing a side chain R which may correspond to a synthetic side chain (including heteroatoms, cyclic or acyclic groups or metal binding groups; also radioisotopic or isotopic substitutions);

iv) amino acids in which the carboxylate is replaced by other groups such as sulfonyl, phosphoryl, phosphonyl and the like; and

v) amino acids with restricted ψ, ϕ angles, such as proline analogues or di-alpha substituted amino acids.

These methods should be applicable to essentially any protein, particularly enzymes displaying catalytic, receptor or ligand binding functions, structural or mixed functionalities. Among the catalytic proteins are included, but not limited to peptidases, nucleases, glycosidases, mono- and dioxygenases, pyridoxalphosphate and flavin dependent enzymes, lipases and aldolases. Among the receptor proteins are included, but not limited to, antibodies, T-cell receptors, muscarinic receptors, G-proteins, lectins, DNA binding proteins and cytochromes. Among the structural proteins are included, but not limited to myosin and silk.

The term "substrate binding site" refers to those the portions of the polypeptide chain whose amino acids are located near to (or are important in conferring) the native three-dimensional spatial conformation of the protein important in substrate binding, or those amino acids situated nearby in space to the region where a substrate or ligand is bound to the polypeptide backbone.

The term "enzymatic active site" refers to those amino acid residues which are situated near to or are involved in conferring the essential spatial or chemical properties necessary for an enzyme to catalyze a reaction.

The term "protein-protein interface" refers to the residues nearby the region where distinct polypeptide chains interact.

The term "cofactor binding site" refers to those residues involved in, or nearby the site where a cofactor or ligand becomes attached or are involved in the recognition for where such might be attached.

It will be recognized that there are a multitude of ways to generate nucleic acid sequences containing the selected codon at the specific site and which will translate to create the polypeptide sequence of interest. These methods include but are not limited to use of natural sequences, modifications of natural sequences, partially or

wholly synthetic sequences and combinations of various natural sequences to create hybrid new proteins. See, Maniatis; Wu and Grossman, Methods in Enzymology, Vol. 153, and Ausubel et al., (1987) Current Protocols in Molecular Biology, Vols. 1 and 2, each of which is hereby specifically incorporated by reference.

Such sequences would include both DNA forms and RNA forms. DNA forms would include, but are not limited to, sequences integrated into a genome, sequences integrated into extrachromosomal elements (including plasmids, episomes and minichromosomes or other free DNAs), phages, viruses, and other similar forms. Similar RNA forms are also included.

The sequence of the translated RNA may be changed by substituting different redundant codons at various sites. It is not well understood why one specific codon is used instead of another redundant one, and each redundant codon might be replaced with one of them. In theory, in the absence of "wobble", one could generate a translation system which would utilize as many as 63 different amino acids plus one termination codon (see, Watson et al., Gene). By application of these techniques, one could generate a translation system with a genetic code quite different from the "universal code". In particular, the starting sequence for the desired product may be natural, a modified sequence or a totally synthetic one.

The site of the substitution may be changed to any codon which is intended to be generally "dedicated" to insertion of the specified preselected amino acid. In the preferred form of use, one would select a codon which does not participate in a "wobble" type redundancy and thus be subject to being translated by an aminoacyl tRNA different from the one selected to perform the misreading. While a truly unique correspondence between the codon and aminoacyl tRNA would provide virtually stoichiometric substitution, this would be diluted through any mechanism which would allow another aminoacyl tRNA to function as the adapter molecule recognizing the selected mistranslation codon. Thus, although any codon could theoretically be chosen to code for

the mistranslation, one would normally select a codon which is not utilized in the reading frame anywhere else in the polypeptide, and would not be translated with any existing aminoacyl tRNA contained in the ultimate translation system to be used.

For these reasons, the termination codons are well suited because there will not be other in frame termination codons in the sequence. Optimally a termination codon different from that actually used to terminate translation would be selected. Furthermore, the unique codon selected may be unique by virtue of having been made so by gene synthesis. Uniqueness would result from substituting all other sites containing that codon to different redundant codons, thus leaving that particular site as the sole site containing the selected codon. In addition, this system could be easily used to made two or more substitutions, of the same predetermined amino acid analogue, or of two or more different analogues, by virtue of selecting multiple unique codons.

The term "substantially homogeneous" relates to the concept of homogeneity of modifications with respect to both site and type. A particular modification is substantially homogeneous when a large majority of the resulting translation product is homogeneous, typically greater than 60% are of a single form, preferably greater than 80% identical, and optimally virtually all, greater than 98%, are identical.

The term "substantially stoichiometric" refers to the property that most of the products are substituted at an intended site, typically more than about 70 to 80% of the products are substituted, preferably more than 90% are substituted, and optimally virtually all, more than 98%, are substituted.

A "protein synthesizing system" is a system which comprises ribosomes, tRNAs, elongation factors and all of the other components necessary to translate a mRNA into protein upon providing the mRNA and appropriate conditions. It is also referred to as a "translation system". Typically, a

cell inherently possesses a protein synthesizing system, but which may have a low level of activity for various reasons. While in vivo systems may be utilized, for the uses described herein, difficulties associated with the introduction of necessary aminoacylated tRNA analogues may exist. This may be achieved by standard microinjection procedures or by any other mechanism of introducing externally produced molecules into the cell, such as electroporation of spheroplasts. An obvious possible technique is either cell or liposome fusions, using such procedures as polyethylene glycol or Sendai viral fusions.

One preferred translation system is the frog oocyte with microinjection, which will also find use for translation systems in other large cells. More typically, an in vitro system is preferred because it is easier to introduce a higher concentration of charged unnatural aminoacyl tRNA molecules to the system. Such systems are available commercially and have been derived from lysates of cells from E. coli, S. cerevisiae, wheat germ and rabbit reticulocytes.

Inherent in the procedure is the capability for using the same single message to direct different translation systems which incorporate distinct unusual aminoacyl tRNAs. Different translation systems may be utilized to incorporate a different unusual amino acid into the selected site, thus generating a series of products, each differing by the insertion of the appropriate preselected unusual amino acid at the selected site. This also makes a termination codon the preferred selected codon as it would not otherwise have a corresponding aminoacyl tRNA. The proper translation systems can be made without the need to remove a preexisting tRNA, but merely by the addition of a new one.

Synthesis of a functional unusual aminoacyl tRNA involves a complicated process of:

a) selection of the correct anticodon to use; and

b) attachment of the appropriate predetermined amino acid analogue.

Once the mistranslation codon is selected, a tRNA with the corresponding anticodon must either be selected or

manufactured. Selection is preferably performed by isolating a natural tRNA. Manufacture may be by mutation and selection, or by site specifically introducing the appropriate anticodon.

The modified translation system normally will not utilize the enzymatic acylation of the unusual adapter molecule, thus the functional definition of tRNA need not normally include the enzymatic acylation function. This, however, does not preclude the use of enzymatic acylation where possible, in which case an acylation capability would be important.

The attachment of the amino acid to a tRNA is naturally catalyzed by the aminoacyl tRNA synthetases. These enzymes are reversible and are highly specific both for the appropriate tRNA (though acylation specificity does not use the anticodon for recognition) and for the amino acid to be attached. Although it may occasionally be possible to use the natural synthetases, or perhaps to modify their specificity, such will be unusual. Thus, the synthesis of the appropriate aminoacyl tRNA is very important. And because the reaction is reversible, it is important that the unusual aminoacyl tRNA not be subject to deacylation activity, either due to the aminoacyl tRNA's inherent inability to act as a substrate for any synthetase present, or by elimination of the deacylating synthetase by some method (inactivation by mutation or chemical modification, through antibody removal or some other means).

The aminoacyl tRNA is of central importance to this invention and the synthesis of the molecule is a major aspect. Where no synthetase exists for an unusual amino acid, some means must be devised to make the adapter molecule.

Infrequently, the unusual amino acid might be a substrate for a synthetase and be charged onto an appropriate tRNA. In some few other cases, it might be possible to modify the amino acid portion of an aminoacyl tRNA without destroying its adapter function. This will seldom be feasible because the occurrence of reactive groups on the

nucleic acid portion will normally compete and eliminate the specificity of the modification reactions.

An alternative method for the synthesis of an unusual aminoacyl tRNA is to synthesize an aminoacyl nucleotide, and then to ligate this moiety onto the appropriate tRNA(-Z) molecule. This method is generally applicable for virtually any aminoacyl tRNA molecule, including attaching normal amino acids, though much less efficient than the synthetase reactions. The only restraints are that the unusual amino acid not interfere with the acylation or deprotection steps and that it not interfere with the ligation step. If so, there is likely to be alternative chemistry to synthesize the adapter molecule. The general scheme is to attach the unusual amino acid onto an oligonucleotide and then to ligate together the nucleotide portions, preferably with T4 RNA ligase. A dinucleotide is preferred because it minimizes interference in the chemistry linking the amino acid to the nucleotides and provides a higher efficiency of ligation than a single nucleotide or AppA analogue. Normally the 3' terminal nucleotides on a tRNA are 5'-pCpCpA-3', so the dinucleotide of choice is 5'-pCpA-3'. It would likely be possible to use other nucleotides (either di- or oligo) such as deoxy-C-ribo-A (i.e., pdCpA) or an entire deoxy-RNA (i.e., DNA) with a 3' terminal ribo-A. After attachment of the unusual amino acid to the nucleotides, preferably using a dinucleotide, the tRNA(-Z) is ligated to the aminoacyl-multi-nucleotide (aa-MNM) to generate the final aminoacyl tRNA analogue.

The ligation step is performed by chemistry or by enzymatic means, the enzyme may be any which has ligation activity on single stranded RNA molecules. The dinucleotide is a sufficiently long substrate for the T4 RNA ligase used in the examples, other enzymes might require a longer or shorter substrate. It will also be observed that the deprotection reactions might, in some cases, be performed after the ligation step.

The source of the tRNA(-Z) component may come from processed natural tRNAs. One source is gene synthesis of a

$tRNA^{Gly}_{CUA}$, where the change of the anticodon of natural $tRNA^{Gly}$ to a termination suppressor destroys recognition of the tRNA by the aminoacyl tRNA synthetases.

Another alternative includes making suppressor tRNA(-Z)'s by "runoff transcription", which will not be modified as normal tRNAs, but having some substantial fraction of activity in translation. Since each tRNA molecule will typically be acylated chemically only once (as opposed to the normal enzymatic reaction), it is generally preferable to create an aminoacyl tRNA that may function somewhat less efficiently in the elongation reaction, if very large quantities of the appropriate tRNA(-Z) can be made for convenient performance of the acylation chemistry. Use of specially designed systems for transcribing the appropriate tRNA acceptor molecules at high efficiency, but not modified, may be very important and are included as possible sources of these molecules. Thus, unmodified tRNA molecules are included in the specifications even though not included in the normal definition of tRNAs.

The chemical procedure of making the aminoacyl tRNAs may be easily modified from the described method. The most obvious is to use a slightly modified tRNA(-Z), which may be slightly longer or shorter or modified from the starting molecule. These molecules are hereby included expressly in the specifications. Another obvious modification is to use, instead of a dinucleotide, a mononucleotide, trinucleotide, or other oligonucleotide. These are also included in the specifications, all included in the multi-nucleotide (MNM) molecule definition. Use of a different nucleotide combination from 5'-pCpA-3' is a likely possibility.

There are at least three acylation routes which may be used to synthesize the aminoacyl-dinucleotides (see Figure 1). Generally, they include blocking particularly reactive groups on the dinucleotides, attachment of the amino acid to the 3' terminal ribose ring and then removal of the blocking groups.

The first route involves treatment of the dinucleotides with nitrophenylsulfenyl chloride (NPS-Cl) to block the cytidine base group. Reaction with an aminoacyl-NPS in 1,1'-carbonyldiimidazole (CDI) will attach the derivatized amino acid to hydroxyl groups on the 3' nucleotide ribose ring. Treatment with thiosulfate will remove the NPS from the cytidine leaving the aminoacyl-dinucleotide.

A second route involves direct synthesis of a dinucleotide or treatment of the dinucleotide with 9-fluorenylmethyloxycarbonyl chloride (FMOCCl) β-cyanoethyl chloride (EtCNOCl) and tetrahydropyranyl chloride (THPCl), which will block the phosphoryl groups, nucleotide base and ribose 2' hydroxyl groups. Reaction with aminoacyl - 2-(4-biphenyl) isopropyloxycarboxyl in CDI will attach to the 3' hydroxyl group. Treatment with 1,1,3,3 tetramethylguanidine, 2-pyridinealdoxime and formic acid will remove all the blocking groups to yield the aminoacyl-dinucleotide. An aminoacyl - vinyloxycarbonyl (aa-VOC) may be substituted for the aminoacyl-BPOC.

A third method of synthesis involves synthesis of or treatment of the dinucleotide with benzyloxycarbonyl (CBZ) and tetrahydropyranyl (THP) resulting in blocking of the cytidine base and the ribose 2' and 3' hydroxyl groups. Reaction with aminoacyl - benzyloxycarbonyl in CDI will cause attachment to the ribose 2'OH group. Treatment with palladium and $BaSO_4$ in H_2 and acid will remove the blocking groups to yield the aminoacyl-dinucleotide. It has been demonstrated that the carbobenzoxy amino acids can be coupled to $pC^{NPS}pA$ and the NPS and CBZ groups removed in 35% overall yield.

There may also be enzymes with the present capability to acylate nucleotides or may soon be engineered to be able to acylate the oligonucleotides or tRNA (Thirsod and Klibanov, (1986) J. Am. Chem. Soc., 108:5638 & 3977; Sweers and Wong, (1986) J. Am. Chem. Soc., 108:6421; Shaw and Klibanov, (1987) Biotech. Bioeng., 29:648; and Wong et al.,

Fluorocarbohydrates: Chemistry & Biochemistry, Taylor (Ed.), ACS Symposium Series, ACS Washington, D.C.).

The selection of a predetermined amino acid analogue to incorporate into the polypeptide will be driven by the needs of the user. Some may desire to substitute any of a number of specific modified amino acids into the site, for any of a number of different purposes. In particular, those of most interest will be those residues which may modify specificity, activity or structure of the protein to satisfy new requirements. Part of the power of this technique is the important potential to break out of the previous limitation of choices among only the natural amino acids. The present invention allows substitution of virtually any L-amino acid, natural or unnatural, as well as D-amino acids.

Among various uses for the proteins of the present invention are the introduction of particular types of residues including, but not limited to, incorporation of:

a) heavy metal atoms (useful in crystallography);

b) cross linking agents;

c) markers (such as radioactive, spectroscopic, fluorescent, magnetic, and electronic);

d) electron acceptors or donors;

e) metal chelators;

f) structurally restricting residues; and

g) residues with novel nucleophilicites;

h) residues with altered acidities and basicites;

i) residues with altered geometries (such as homoserine, homocysteine or ornithine); and

j) residues with altered hydrogen bonding properties (e.g., amidine vs. amide).

The physical or biochemical properties which can be studied by making substitutions at known and specific sites are many. Spectroscopic markers may be introduced to particular regions in the tertiary structure of a protein or complex of polypeptides. Residues may be introduced with a different pKa, or which will affect the pKa of nearby residues, with a different nucleophillicity, or which will

affect the nucleophillicity of nearby residues, with electron acceptor function, with metal chelator function, with modified hydrogen bond donor or acceptor function, with altered or restricted bond torsion angles, with cofactor binding capability or with special markers for fluorescence or other detection or purification methods. The invention provides the opportunity to introduce into the polypeptide chemical groups which are beyond the range of the natural amino acid residues, and to escape from many of the constraints previously imposed by nature.

In particular, one of the most important uses will be the incorporation of heavy metal scattering centers into identical locations in a polypeptide, which will, upon crystallization, allow for relative ease in solving of the wave equations necessary to determine the gross three-dimensional structure of a polypeptide chain. The structure of a protein is very important, and is normally the essential property of an enzyme which confers on it the ability to perform its function. These functions will include aspects of the properties of mechanism of catalysis, specificity of substrate binding and reaction, structural features and regulatory interactions.

Although many of the techniques may be most suitable for evaluating the physical or biochemical properties within some short distance away from the portions of the polypeptide chain localized nearby in space to substrate binding sites, enzymatic active sites, ligand binding sites, protein-protein interfaces and cofactor binding sites. Nearby, as used herein, means within about 50 to 150 angstroms from the site of interest, preferably less than about 30 angstroms, and optimally within 15 angstroms in three dimensional space.

In vitro translation systems will normally be used herein, though in some cases it may be possible to use in vivo systems. Typical in vitro translation systems are procaryote sources including E. coli, and eucaryote sources including rabbit reticulocyte lysates, wheat germ lysates and yeast lysates and heterogenous mixed systems containing

components from various sources. (See, Wu and Grossman, Methods in Enzymology, Vol. 153). Preferred translation systems include modified transcription and translation systems exhibiting greatly increased transcription by placing the gene of interest under control of an operably linked strong promoter. Alternatively, incorporation of a very active RNA polymerase would increase the message level. Background products might be lowered by using a rifampicin insensitive T7 RNA polymerase to transcribe the message operably linked to a T7 specific promoter, while other endogenous transcripts are repressed by the presence of rifampicin. A continuous flow translation system is a possible improvement for large scale production purposes (Spirin et al., (1988) Science 242:1162). Intracellular injection into frog oocytes, muscle cells or other large cells might be performed to introduce the message, tRNA or aminoacyl tRNA into the cells.

In certain cases, one particular translation system source would be preferred. The yield of desired product or further processing may be dependent upon the presence or absence of activities in the translation systems. Such might include glycosylation, acetylation or other processing enzymes, or lack of proteases or other enzymes.

Among the important inherent advantages of this method of analysis is the ability to make numerous substitutions at each selected site. Upon construction of a single mRNA having a mistranslation codon, a virtually limitless number of different substitutions may be made at that site limited only by the number of unnatural aminoacyl tRNA molecules desired to be utilized in the translation and mistranslation process. The selection of a termination codon is especially useful because no endogenous tRNA need be removed or inactivated. Inherently, no natural tRNA would be present, and the new unnatural amino acid analogue is incorporated merely by addition of the new aminoacyl tRNA analogue. With selection of the appropriate tRNA from a catalogue of unnatural aminoacyl tRNAs, any unnatural amino acid analogue could be substituted at the position selected.

Among the unnatural amino acids which may be selected are modifications of the natural amino acids, modifications of amino acids other than the natural ones, amino acids other than alpha-amino acids (i.e. beta, gamma, etc.), amino acids having a different stereospecificity (i.e. D- amino acids, or having a different stereospecificity at other asymmetric carbon or other atoms), amino acids having substituted atoms or containing unusual elements and residues containing cofactor binding sites.

A coupled transcription and translation system is one in which the products of the transcription system are directly translated by the system without purification or isolation of the mRNA produced. The system is initially run under conditions which are optimum for transcriptional activity after which the conditions are optimized for translation of the transcripts produced.

It is possible to specifically design a translation system which provides the essential features of this system, while using a translation system of significantly modified code. For this reason, a translation system having a totally different correspondence between codon and amino acid is also included.

The protein products of this method may have a variety of properties, such as a) homogeneity of site and type of modifications in the proteins; b) stoichmetric modification (all of the subject proteins are modified, without dilution by unmodified forms; and c) known characterization for type and position of the modification.

The large number of potential uses of such products should be immediately recognized by any protein biochemist. Any means of characterizing a protein should be simplified by a lowered background or noise from unmodified forms or heterogeneously modified forms. The means for physical or biochemical analysis is as broad as the techniques available and applied to purified proteins, see, for example, Methods in Enzymology, Vols. 1-187; Lehninger, Biochemistry; Stryer Biochemistry; and Creighton, The Proteins.

For example, in applying the technique of characterization by X-ray crystallography, the introduction of a heavy metal scattering center in a unique and uniform site in a protein will greatly assist in the analysis of the wave pattern data to solve the wave equations necessary to determine the three dimensional protein crystal structure (Mathews, (1976) Annual Review of Physical Chemistry, 27:493-523).

In spectroscopy applications, introduction of selected particular physical markers at unique and uniform sites in the protein should be extremely useful in characterizing the static and dynamic fine chemical and electronic structure of binding sites, active sites and other important portions of a protein

The following experimental section is offered by way of example and not by limitation.

EXPERIMENTAL

Mutagenesis and In Vitro Protein Synthesis

The mutagenesis methodology described above has been developed with the well-characterized hydrolytic enzyme, RTEM β-lactamase (Hamilton-Miller and Smith Eds., (1979) "Beta-Lactamases", Academic Press, London; Jack and Sykes, (1971) Ann. N.Y. Acad. Sci. 182:243; and Abraham, (1977) J. Antibiot. 30:51; Fisher et al., (1978) Biochemistry 17:2180). This bacterial (E. coli) enzyme is a single chain 29kD protein containing one disulfide bond (Sutcliffe, (1978) Proc. Natl. Acad. Sci USA 75:3737; Ambler and Scott, (1978) Proc. Natl. Acad. Sci. USA 75:3732; and Pollitt and Zalkin, (1983) J. Bacteriol. 153:27). The gene encoding β-lactamase has been sequenced (Sutcliffe, (1978) Proc. Natl. Acad. Sci USA 75:3737; Ambler and Scott, (1978) Proc. Natl. Acad. Sci. USA 75:3732; and Pollitt and Zalkin, (1983) J. Bacteriol. 153:27), the three dimensional structure of a homologous Class A β-lactamase has been solved (Herzberg and Moult, (1987) Science 236:694) and a simple spectrophotometric assay exists for enzyme activity (One nitrocefin hydrolysis unit (1μmole nitrocefin hydrolyzed/min/mL, 0.1 mM nitrocefin, 50mM phosphate buffer, pH 7) corresponds to 0.61 μg enzyme, as determined by Bradford assay; O'Callaghan et al., (1972) Antimicrob. Ag. Chemother. 1:283). The enzyme inactivates β-lactam antibiotics (penicillins and cephalosporins) by hydrolyzing the β-lactam amide bond. The reaction proceeds via a two step mechanism involving nucleophilic attack of Ser70 to form an acyl-enzyme intermediate, which is then hydrolyzed to yield the corresponding acid and free enzyme (Fisher et al., (1980) Biochemistry 19:2895; Knowles, (1985) Acc. Chem. Res. 18:97; Dalbadie-McFarland et al., (1982) Proc. Natl. Acad. Sci. USA 79:6409; and Sigal et al., (1984) J. Biol. Chem. 259:5327).

Phe66, which is conserved in 4 Class A β-lactamases (Ambler, (1979) "Beta-Lactamases", Hamilton-Miller and Smith Eds., Academic Press, New York pp. 99-125), was chosen as the first target for mutagenesis since a number of L-phenylalanine analogues are easily synthesized and

phenylalanine does not require additional side chain protection in the chemical aminoacylation step. A 2.5 Å crystal structure of the S. aureus enzyme (33% homology with the E. coli enzyme) localizes the residue to an extended loop between a buried β-sheet and an α-helical domain containing the active site (Herzberg and Moult, (1987) Science 236:694). The structural importance of this residue was confirmed by constructing the Phe66 → Ala (pF66A) and Phe66 → Tyr (pF66Y) mutants (Fig. 3), both of which yielded little activity in crude cell extracts. (All in vivo work was carried out using E. coli strain JM101 (*Δlacpro thi, supE, F'traD36, proAB, lacI^q ZΔM15*). [C.Janisch-Perron et al., (1983) Gene 22:103]. All in vitro work was done using S-30 extracts [Pratt, (1984) "Transcription and Translation", Hans and Higgins, (Eds.), IRL Press, Oxford] prepared from E. coli strain D10 (*rna-10, relA1, spoT1, metB1*) [Gesteland (1966) J. Mol. Biol. 16:67]). Attempts at purification resulted in loss of all activity for the F66A mutant, while the F66Y mutant was purified in low yield and characterized. The K_M of the F66Y mutant for nitrocefin was identical to that of the wild-type enzyme, whereas the k_{cat} was 16% that of wild-type enzyme (results not shown).

In vivo and in vitro synthesis of β-lactamase was carried out using the plasmid pSG7 (Fig. 3), which was designed with the following considerations: RTEM β-lactamase is synthesized in vivo with a 23-amino acid leader sequence that is clipped off during translocation across the inner membrane to yield fully active enzyme. In order to express active enzyme in vitro, we used a truncated gene for β-lactamase (Kadonaga et al., (1984) J. Biol. Chem. 259:2149) in which a 63-bp deletion corresponding to a 21-amino acid deletion in the leader sequence is sufficient for direct expression of active enzyme. The truncated gene was placed under the transcriptional control of the strong hybrid tac promoter (Amann et al., (1983) Gene 25:167), as it has been demonstrated that the amount of protein synthesized in an in vitro translation system is proportional to the amount of mRNA added (Reiness and Zubay, (1973) Biochem. Biophys. Res.

Comm. 53:967). To this end, the truncated gene was also placed under control of the Φ promoter (Tabor and Richardson, (1985) Proc. Natl. Acad. Sci. USA 82:1974) from bacteriophage T7 with the intent of supplementing the reaction with T7 RNA polymerase, which synthesizes RNA at a rate 10 times that of the E. coli polymerase (Chamberlin and Ryan, (1982) The Enzymes 15:87). The kanamycin resistance gene from Tn903 (Oka et al., (1981) J. Mol. Biol. 147:217), cloned in the opposite orientation from the tac and T7 promoters, provides a selectable marker for these plasmids.

Protein synthesis was carried out in vitro in order to simplify addition of the aminoacylated suppressor tRNA to the translational machinery. The coupled E. coli system developed by Zubay (Zubay, (1973) Annu. Rev. Gen. 7:267), with some modifications by Collins (Collins, (1979) Gene 6:29) and Pratt (Pratt, (1984) "Transcription and Translation" Hanes and Higgins (Eds.), IRS Press, Oxford, pp.179-209), was used with little further modification except for lowering the pH of the system from 8.2 to 7.4 in order to better stabilize the base-labile acyl linkage of the added aminoacylated suppressor (Fig. 4).

Yields of active β-lactamase synthesized in this system primed with pSG7 typically ranged from 30-45 μg/mL of reaction mixture, based on the nitrocefin hydrolysis assay (see Fig. 6). This corresponds to 23-33 copies of active enzyme per copy of gene, and represents an 11-fold increase in synthesized enzyme over that directed from the wild-type Apr promoter of the pBR322 derivative pSG1. (The amount of overproduction in vivo, that is, JM101/pSG7 vs. JM101/pSG1, is also 11-fold, based on the specific activity of crude cell extracts.) Surprisingly, addition of T7 RNA polymerase (to a final concentration of 8500 units/mL) to reactions primed with the T7 promoter plasmid pSG1 yielded levels of active enzyme that were 65-70% of the levels produced in reactions primed with pSG7. (In vitro protein synthesis of 434 repressor expressed behind the strong tac promoter affords greater than 150 copies of protein per copy of gene.) In vitro produced β-lactamase was purified to

homogeneity by ammonium sulfate precipitation followed by chromatofocusing and anion exchange chromatography (Fig. 4). Protein was determined to be homogeneous by SDS-polyacrylamide gel electrophoresis and had a k_{cat} and K_M for nitrocefin identical to that of in vivo produced enzyme. All suppression work was carried out using the pSG7 derivative pF66am (Fig. 3), which carries the Phe66 → TAG mutation.

Suppressor tRNA Generation Characterization

The suppressor tRNA used to deliver the unique amino acid to the growing peptide chain on the ribosome must meet two criteria: it must efficiently insert the amino acid in response to the UAG message and it must be neither acylated nor deacylated by any of the E. coli aminoacyl-tRNA synthetases present in the in vitro transcription/translation system. The first condition is necessary for producing quantities of protein that can be purified and further studied, the second condition is required to insure that only the desired unnatural amino acid and not one or more of the twenty natural amino acids in the in vitro reaction will be inserted into the protein (Schimmel and Soll (1979) Ann. Rev. Biochem. 48:601; Fersht and Kaethner, (1976) Biochemistry 15:3342; Igoli et al., (1978) Biochemistry 17:3459; Schreier and Schimmel (1972) Biochemistry 11:1582; and Yarus, (1972) Proc. Natl. Acad. Sci. USA 69:1915). An amber suppressor tRNA derived from yeast $tRNA^{Phe}$ (Bruce and Uhlenbeck (1982) Biochemistry 21:3921 and 21:855) was expected to meet these requirements based on the following observations: Yeast $tRNA^{Phe}_{CUA}$, in which residues 34-37 of yeast $tRNA^{Phe}$ are replaced by 5'- CUAA-3', is expected to be an efficient suppressor based on Yarus' extended anticodon loop hypothesis (Yarus, (1982) Science 218:646). In addition, Bruce and coworkers (Miller et al., (1977) J. Mol. Bio. 109:275; Bruce et al., (1982) Proc. Natl. Acad. Sci.USA 79:7127; Bossi and Roth, (1980) Nature 286:123; and Steege, (1978) "Biological Regulation and Development", Vol. I, Goldberg Ed., Plenum Press) demonstrated that yeast $tRNA^{Phe}_{CUA}$ was efficient in translating UAG codons in a mammalian protein synthesizing

system (although being somewhat less efficient in a wheat germ system). Kwok and coworkers (Kwok and Wong, (1980) Can. J. Biochem. 58:213) have shown that E. coli phenylalanyl-tRNA synthetase (PRS) aminoacylates yeast $tRNA^{Phe}$ ≤ 1% as well as it acylates E. coli $tRNA^{Phe}$.

Yeast $tRNA^{Phe}_{CUA}$ was prepared in milligram quantities according to the anticodon-loop replacement procedure of Bruce and Uhlenbeck (Bruce and Uhlenbeck, (1982) Biochemistry 21:855 and 21:3921) (Fig. 5). This procedure involves removal of the three anticodon nucleotides G-34, A-35 and A-36 as well as the modified nucleotide Y-37 from the anticodon loop of yeast $tRNA^{Phe}$. The four excised nucleotides are then replaced with a chemically synthesized CpUpApA which includes the anticodon sequence required for an amber suppressor tRNA.

The details of this modified procedure of Bruce and Uhenbeck (Bruce and Uhlenbeck, (1982) Biochemistry 21:855 and 21:3921) are as follows. After the initial depurination step the $tRNA-_Y$ was recovered by ethanol precipitation and treated with aniline hydrochloride. The cleavage products were recovered by EtOH precipitation and separated on a preparative denaturing PAGE gel (8%, 0.015cm x 16cm x 42cm; 5 mg of the crude tRNA was loaded onto each gel). The bands were stained with 0.02% toluidine blue, cut out and eluted with 2 x 10mL 100mM NaOAc (pH 4.5), 1mM EDTA and 0.1% SDS. The stain was removed by extractions with phenol and $CHCl_3$, the tRNA half-molecules were recovered by ethanol precipitation. For the partial nuclease digestion the concentration of the RNase A was increased to 2 μg/mL, and following ethanol precipitation, the pellet was resuspended in sterile water and extracted with phenol, phenol:$CHCl_3$ (1:1), $CHCl_3$ and reprecipitated with ethanol. The ribonucleotide tetramer CpUpApA was synthesized by sequential phosphotriester coupling of protected nucleosides (Jones et al., (1980) Tetrahedron 36:3015; Boom and Wreesman, (1984) "Oligonucleotide Synthesis", Gait Ed., IRL Press, Washington). The fully deprotected CUAA was ligated onto the 3' tRNA half-molecule using T4 RNA ligase supplied by Takara

Shuzo. The conditions for the ligation were 50 μM ATP, 19 μM tRNA (both half-molecule s are present in the reaction), 820 μM CUAA and 50 U/mL T4 RNA ligase. Ligation reactions were typically carried out on 5 mg of the RNase A-digested tRNA in a reaction volume of 10 mL. The kinase treatment was carried out with 4 μM tRNA, 120 μM ATP and 50 U/mL of T4 polynucleotide kinase (Richardson (1965) Proc. Natl. Acad. Sci. USA 54:158; and Midgley and Murray (1985) EMBO J. 4:2695). The final ligation was done with 25 U/mL of T4 RNA ligase.

The suppressor produced by this method is missing the 3' terminal pCpA aminoacyl acceptor stem. These nucleotides can be replaced using the tRNA repair enzyme nucleotidyl transferase (Cudny and Deutscher, (1986) J. Biol. Chem. 261:6450) to yield a full-length yeast $tRNA^{Phe}_{CUA}$. The suppressor tRNA can be aminoacylated in vitro with [^{3}H]-Phe to levels of 30-35% (based on radioactivity incorporated into purified [^{3}H]-Phe - $tRNA^{Phe}_{CUA}$) using a large excess of yeast PRS. Enzymatic misacylation reactions (300 μL total volume) contained the following: 4 μM $tRNA^{Phe}_{CUA}$ (30μg, which had been desalted and lyophilized following gel purification), 80 μM phenylalanine, 40 mM Tris-HCl (pH 8.5), 15 mM $MgCl_2$, 45 μg/mL BSA, 3.3 mM DTT, 2 mM ATP and 22 Units yeast PRS (where 1 unit activity incorporates 100 pmol Phe in 2 minutes at 37°C under the following conditions: 2 μM $tRNA^{Phe}$ (Boehringer Mannheim), 2 mM ATP, 3.3 mM DTT, 8.1 μM Phe, 40 mM Na HEPES (pH 7.4), 15 mM $MgCl_2$, 25 mM KCl, and 50 μg/mL BSA). The reaction mixture was incubated at 37°C for 3 minutes, then quenched by addition of 2.5 M NaOAc (pH 4.5) to 10% v/v. The quenched reaction was immediately extracted with phenol (pre-equilibrated with 0.25 M NaOAc, pH 4.5), phenol:$CHCl_3$ (1:1), $CHCl_3$, then precipitated with EtOH. The extraction and precipitation were repeated twice more. The tRNA was then desalted on a Pharmacia fast desalting column and lyophilized. the lyophilized mixture of acylated and non-acylated tRNA was stored at -80°C until immediately prior to its use in in vitro protein synthesis reactions.

Under similar reaction conditions wild-type yeast $tRNA^{Phe}$ acylates to levels of 40-45% with yeast PRS. Attempts at separating acylated from non-acylated tRNA's by BD-cellulose chromatography (Kreig et al., (1986) Proc. Natl. Acad. Sci. USA 83:8604; Wiedmann et al., (1987) Nature (London) 328:830; Johnson et al., (1976) Biochemistry 15:569; Baldini, et al., (1988) Biochemistry 27:7951; Heckler et al., (1984) Tetrahedron 40:87; and Heckler et al., (1984) Biochemistry 23:1468) resulted in poor separation and unacceptably low yields of acylated tRNA. The mixture of acylated and non-acylated tRNA's is therefore used directly in in vitro protein synthesis reactions. We have attempted to misacylate $tRNA^{Phe}_{CUA}$ with several analogues of phenylalanine using yeast PRS, however these experiments were unsuccessful under variations in pH, concentration of buffer and/or salt, and concentration of organic solvents.

Importantly, yeast $tRNA^{Phe}_{CUA}$ is not recognized by the E. coli aminoacyl-tRNA synthetases present in our in vitro system (Fig. 6). An in vitro reaction primed with pF66am and non-acylated suppressor, in the presence of [^{3}H] phenylalanine, results in no β-lactamase activity and no radioactive band of the correct molecular weight when analyzed on a denaturing polyacrylamide gel. A reaction primed with pF66am and [^{3}H]-Phe-$tRNA^{Phe}_{CUA}$, the level of in vitro β-lactamase synthesis from pF66am is 15-20% compared to that for pSG7. These results demonstrate that yeast $tRNA^{Phe}_{CUA}$ meets the design criteria outlined above: the tRNA is not enzymatically aminoacylated or deacylated, and it inserts an acylated amino acid in response to UAG.

A run-off transcript tRNA was produced (Sampson and Uhlenbeck, (1988) Proc. Natl. Acad. Sci. USA 85:1033) corresponding to the sequence of yeast $tRNA^{Phe}_{CUA}$ (-CA) which has been purified and ligated to phenylalanyl-pCpA (Suich and Noren, unpublished results). Addition of this aminoacylated tRNA to in vitro protein synthesis reactions affords 50-65% of the β-lactamase activity observed for reactions using the same amounts of Phe-$tRNA^{Phe}_{CUA}$ derived from anticodon loop replacement. This level of translational efficiency for the

run-off suppressor compares favorably to data for a $tRNA^{Gly}$ constructed in a similar fashion (Samuelsson et al., (1988) J. Biol. Chem. 263:1392).

Chemical Aminoacylation

As noted above, enzymatic misacylation by the aminoacyl-tRNA synthetases is not a general method due to the high specificity of these enzymes. Chemical misacylation, however, should be generalizable to any amino acid-like structure. Direct chemical acylation of an intact tRNA is not practical due to the large number of reactive sites in the macromolecule. Hecht and coworkers (Heckler et al., (1984) Tetrahedron 40:87; and Heckler et al., (1984) Biochemistry 23:1468) simplified this problem by chemically acylating the dinucleotide pCpA and enzymatically ligating it to the 3' terminus of a truncated tRNA [tRNA(-CA)] using T4 RNA ligase to afford an aminoacyl-tRNA. This approach, though successful, suffered two major drawbacks: the α-amino protecting group was not removed, restricting the aminoacyl-tRNA to only act as a P site donor, and the chemical acylation yield was quite low.

The general strategy for chemical acylation of pCpA involves carboxyl activation of an N-blocked amino acid followed by coupling via an ester linkage to the diol of the terminal adenosine (the 2' and 3' acyl groups rapidly interconvert in aqueous solution). Aminoacylation is complicated by preferential acylation of the exocylic amino group of cytidine and 2', 3' diacylation of adenosine. The α-amino protecting group greatly increases the stability of the aminoacyl ester linkage to hydrolysis and avoids polymerization during carboxyl activation Schubert and Pinck, (1974) Biochimie 56:383). However, the protecting group must be removed if the acylated-tRNA is to function as an A site donor. It has recently been shown by Brunner (Kreig et al., (1986) Proc. Natl. Acad. Sci. USA 83:8604; Wiedmann et al., (1987) Nature (London) 328:830; Johnson et al., (1976) Biochemistry 15:569; Baldini, et al., (1988) Biochemistry 27:7951) that α-amino protected aminoacyl pCpA can be

deprotected before ligation to tRNA(-CA) without hydrolysis of the aminoacyl ester linkage.

The scheme for aminoacylation of $tRNA^{Phe}_{CUA}$ is outlined in Figure 7. A minimal protection scheme was used in which only the exocyclic amine of cytidine was protected by 0-nitrophenyl sulfenyl chloride (NPS-Cl). The α-amino group of the amino acid was also protected with the NPS group. NPS-pCpA was acylated with N-blocked Phe using N,N'-carbonyldiimidazole as the activating agent. The NPS protecting groups were removed from cytidine and the amino acid in high yield using aqueous thiosulfate (Lapidot et al., (1970) Biochem. Biophys. Res. Comm. 38:559 and Heikkilä et al., (1983) Acta. Chem. Scand. B37:857). The acylation/deprotection was carried out in 14% overall yield (unpublished results indicate that 2', 3'-diacylation is the major factor limiting acylation yields in model compounds), which compares favorably with the 3-4% yields of Hecht and Brunner (Kreig et al., (1986) Proc. Natl. Acad. Sci. USA 83:8604; Wiedmann et al., (1987) Nature (London) 328:830; Johnson et al., (1976) Biochemistry 15:569; Baldini, et al., (1988) Biochemistry 27:7951; Heckler et al., (1984) Tetrahedron 40:87; and Heckler et al., (1984) Biochemistry 23:1468). However, Chladek (Happ et al., (1987) J. Org. Chem. 52:5387) has recently reported aminoacylation of 5'-CpCpA in 26% overall yield via an alternate strategy.

Chemical acylation reactions (80μ total volume) contained the following: 600μM pCpA-Phe (40μg), 10 μM $tRNA^{PheUCA}$ (20μg, which had been desalted and lyophilized following gel purification), 55 mM HEPES (pH 7.5), 250μM ATP, 15mM $MgCl_2$, 20mg/mL BSA, DMSO (to 10% v/v) and 200 units T4 RNA ligase. The reaction mixture was incubated at 37°C for 12 minutes, quenched by addition of 2.5 M NaOAc (pH4.5) to 10% v/v and treated as described above, but with only one round of extraction/precipitation. The di-NPS protected aminoacyl pCpA was also a substrate for T4 RNA ligase, but better yields of the aminoacyl tRNA were obtained by deprotection followed by ligation rather than ligation and subsequent deprotection.

Fully deprotected pCpA-Phe was ligated directly to $tRNA^{Phe}_{CUA}$(-CA) using T4 RNA ligase (note that the truncated suppressor tRNA is generated directly by the anticodon loop replacement method). The yield of Phe-$tRNA^{Phe}_{CUA}$ is 35% based on analysis of 3H-Phe incorporation into the purified suppressor (gel electrophoresis indicates 80-90% of the $tRNA^{Phe}_{CUA}$ (-CA) is converted to material with the same mobility of $tRNA^{Phe}_{CUA}$). Using this procedure $tRNA^{Phe}_{CUA}$ (-CA) was also aminoacylated with D-phenylalanine (D-Phe), (S)-p-nitrophenylalanine (p-NO_2-Phe), (S)-homophenylalanine (2-amino-4-phenylbutanoic acid, HPhe), (S)-p-fluorophenylalanine (p-FPhe), (S)-3-amino-2-benzylpropionic acid (ABPA) and (S)-2-hydroxy-3-phenylpropionic acid (PLA) (in this case no α-hydroxyl protection was used). These aminoacyl tRNA's were used in in vitro protein synthesis to synthesize mutant β-lactamases (vide infra). Current efforts to optimize aminoacylation include the use of acid labile protecting groups and protecting groups that can be removed by hydrogenation, as well as an investigation of the use of non-selective lipases for the aminoacylation of unprotected RNA. Protecting groups which can be removed by hydrogenolysis or acid treatment will also simplify protection of unnatural amino acid side chains.

In Vitro Suppression

In vitro reactions primed with pF66am and supplemented with suppressor that had been enzymatically acylated with [^{3}H]-phenylalanine (30% acylated), to a final concentration of 167 μg/mL, yielded 5.5-7.5 μg/mL of active β-lactamase, which represents 15-20% suppression efficiency (Fig. 8). Suppressor that had been chemically acylated using pCpA-Phe (35% acylated) resulted in a yield of 2.8-7.5 μg/mL. This is a sufficient yield for purification of the enzyme to near homogeneity from a 1 mL reaction in 7-15% overall yield (Fig. 8) using ammonium sulfate precipitation followed by chromatofocusing and anion exchange chromatography. Importantly, the k_{cat} and K_M for the purified β-lactamase were identical to those of the wild-type enzyme produced in

vitro (Figure 10). Site-specific insertion of ^{3}H-phenylalanine by [^{3}H]-Phe-tRNA$^{Phe}_{CUA}$ into β-lactamase was verified by peptide mapping experiments. There are twenty-seven putative trypsin cleavage sites, and five phenylalanine residues in TREM β-lactamase (Sutcliffe, (1978) Proc. Natl. Acad. Sci USA 75:3737; Ambler and Scott (1978) Proc. Natl. Acad. Sci. USA 75:3732; Pollitt and Zalkin, (1983) J. Bacteriol. 153:27). These five phenylalanines are distributed in four of the twenty-eight tryptic fragments, with one eight-residue peptide containing both Phe66 and Phe72. β-lactamase was synthesized in vitro from pSG7 in the presence of added [^{3}H]-phenylalanine. The purified radiolabelled enzyme was digested with trypsin and the fragments were separated by reversed-phase FPLC (Fig. 9). Four discrete radioactive peaks were observed, in agreement with the locations of [^{3}H]-Phe in RTEM β-lactamase (Sutcliffe, (1978) Proc. Natl. Acad. Sci USA 75:3737; Ambler and Scott (1978) Proc. Natl. Acad. Sci. USA 75:3732; Pollitt and Zalkin, (1983) J. Bacteriol. 153:27; Fisher et al., (198) Biochemistry 19:2895 and Knowles (1985) Acc. Chem. Res. 18:97). The peak that elutes in fractions 44 and 45 contains twice as many counts as the other three peaks and is assigned as the peptide that contains F66 and F72. A similar analysis of tryptic peptides derived from β-lactamase synthesized in vitro from pF66am in the presence of added [^{3}H]-Phe-tRNA$^{Phe}_{CUA}$ shows one radioactive peak. The fact that radioactivity elutes in fraction 44 in both the wild-type (pSG7) and suppressed (pSF66am) experiments taken together with the observation that in the wild-type experiments this peak contains twice as much radioactivity as the others, strongly suggests that in the suppressed experiment [^{3}H]-Phe is inserted uniquely at the target site (F66).

The phenylalanine analogues D-Phe, HPhe, ABPA and PLA were each loaded onto suppressor tRNA as described above. In vitro protein synthesis reactions carried out in the presence of [^{35}S]-methionine resulted in similar levels of radioactivity incorporated into trichloracetic acid (TCA)-precipitable material for the p-FPhe, p-NO_2Phe and HPhe

reactions. Kinetic analyses of the β-lactamases synthesized in these reactions demonstrate similar K_M's but different k_{cat}'s (Figure 10). Direct quantitation of the purified p-NO_2Phe and HPhe mutants was impossible, as both mutants lost activity during purification attempts. Consequently, [^{35}S]-methionine incorporation and TCA precipitation were used to quantitate all the mutants for the purpose of direct comparison. Experiments with D-Phe, PLA and ABPA resulted in no detectable β-lactamase activity or protein synthesis.

These results demonstrate that in the case of phenylalanine, analogues that differ in both steric and electronic properties can be substituted into proteins. Replacement of Phe66 by Tyr, which differs from Phe both sterically (4-OH group) and electronically (-NO_2 is a good π-electron donor), leads to an approximate twofold decrease in k_{cat} with little effect on K_M. Similarly, replacement of Phe66 by p-nitrophenylalanine, which again differs both sterically (4-NO_2 group) and electronically (-OH is a good π-electron acceptor), leads to an approximate twofold decrease in k_{cat} with little effect on K_M. On the other hand, replacement of Phe66 by p-fluorophenylalanine, which is both sterically and electronically similar to Phe, leads to a slight increase in k_{cat} with no effect on K_M. Replacement of Phe66 with homophenylalanine, which is electronically identical to Phe but substantially different sterically, leads to an approximate sixfold decrease in k_{cat} and an increase in K_M. Both the HPhe and p-NO_2Phe mutants, which correspond to the greatest steric perturbation, are unstable and presumably unfold and proteolyze during purification attempts.

Attempts to alter the protein backbone by replacing the amide linkage with an ester linkage (PLA), adding an additional methylene group into the backbone (ABPA), or changing the sterochemistry of the α-carbon (D-Phe) lead to no detectable synthesized protein or enzymatic activity. It is not clear at present whether these results stem from impaired protein folding or stability (Phe66 is adjacent to a proline residue) or from inability of these amino acids to

function as ribosomal A site acceptors. It has been reported that PLA is incorporated into the amino terminal position of polyphenylalanine (Herve and Chapevill, (1965) J. Mol. Biol. 13:757) and that N-acetyl-D-phenylalanine functions poorly as a P-site donor in response to a poly-(U) message (Rosesser et al. (1986) Biochemistry 25:6361 and Heckler et al., (1983) J. Biol. Chem. 258:4492). Yamane et al. report that K_{diss} for the D-Tyr-tRNA - EF-Tu ternary complex is 25-fold higher than for the L-Tyr-tRNA - EF - Tu ternary complex (Yamane et al., (1981) Biochemistry 20:7059). The same level of stereochemical selectivity for ternary complex formation between EF-Tu and D-Phe-$tRNA^{Phe}_{CUA}$ would result in substantial hydrolysis of the aminoacyl ester during in vitro reactions as well as poor EF-Tu mediated binding of the aminoacyl tRNA to the ribosome.

Sufficient protein can be purified to characterize the catalytic constants and specificity of the mutants, to carry out limited mechanistic and mapping studies and to probe protein structure with techniques such as ESR and fluorescence spectroscopy. Improvements in in vitro protein synthesis, methods for tRNA generation, and tRNA aminoacylation chemistry will permit production of milligram quantities of mutant proteins via this strategy.

From the foregoing, it will be appreciated that the present invention provides improved means for producing modified proteins. The methods are rapid, simple and universal in utility.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the claims.

WE CLAIM:

1. A method for site specifically incorporating an unnatural amino acid analogue into a protein, said method comprising:

(a) introducing a preselected codon into at least one site in a mRNA sequence encoding the protein; and

(b) translating the mRNA sequence in a protein synthesizing system comprising an aminoacyl tRNA analogue capable of polymerizing the unnatural amino acid analogue into a nascent polypeptide chain at the direction of the preselected codon.

2. A method of Claim 1, wherein the protein synthesizing system is an in vitro protein synthesizing system.

3. A method of Claim 2, wherein the in vitro protein synthesizing system is prepared from:

a) E. coli;

b) S. cerevisiae;

c) wheat germ; or

d) rabbit reticulocyte.

4. A method of Claim 1, wherein the preselected codon is a translation termination codon.

5. A method of Claim 4, wherein the preselected codon is a UAG (amber) codon.

6. A method of Claim 4, wherein the preselected codon is inserted at predetermined sites.

7. A method of Claim 1, wherein the unnatural amino acid analogue is one selected from the group of:

i) modified natural amino acids;
ii) modified uncharged amino acids;
iii) modified acidic amino acids;
iv) modified basic amino acids;
v) non-alpha amino acids;
vi) amino acids with altered ψ, ϕ angles; and
vii) amino acids containing functional groups selected from the group of nitro, amidine, hydroxylamine, quinone, aliphatic, cyclic and unsaturated chemical groups.

8. A method of Claim 1, wherein the aminoacyl tRNA analogue is the only aminoacyl tRNA molecule in the protein synthesizing system capable of recognizing the preselected codon.

9. A method of Claim 1, wherein the preselected codon is introduced into one site of the mRNA sequence encoding the protein.

10. A method of Claim 1, wherein the protein has a molecular weight greater than about ten thousand daltons.

11. A method of Claim 1, wherein the unnatural amino acid analogue is situated within about 100 angstroms of:

a) a substrate binding site;
b) an enzymatic active site;
c) a protein-protein interface;
d) a cofactor binding site; or
e) a ligand (agonist or antagonist) binding site.

12. A protein synthesized by the method of Claim 1, 8 or 10.

13. A protein according to Claim 12, wherein the protein is substantially stoichiometrically substituted at one or more predetermined sites.

14. A protein according to Claim 12, wherein the protein is substantially homogeneous in its substitutions at one or more predetermined sites.

15. A protein according to Claim 13, wherein the protein is substantially homogeneous in its substitutions.

16. A method for determination of physical, chemical or biochemical properties of a protein, such method comprising the steps of:

a) synthesizing a substantially pure protein stoichiometrically substituted at specific sites by the method of Claim 1, 8 or 10; and

b) analyzing the physical or biochemical properties of the protein.

17. A method of Claim 16, wherein the protein is analyzed by X-ray crystallography or NMR.

18. A method of Claim 16, wherein analyzing the physical or biochemical properties of the protein determines:

a) static physical properties of the polypeptide chain;

b) mechanism of action of an enzymatic reaction;

c) specificity of protein binding to ligand;

d) dynamic interaction of amino acid residues of a subject protein with a substrate;

e) folding of the protein; or

f) interaction of the protein with other proteins, with nucleic acids or with sugars.

19. A method of Claim 16, wherein the protein is analyzed within about 100 angstroms of the unnatural amino acid analogue.

20. A method for making multiple alternative substitutions at preselected amino acid positions of a protein comprising;

a) producing one mRNA with mistranslation codons at sites in the mRNA corresponding to the preselected amino acid positions; and

b) translating the mRNA in a series of two or more translation systems each comprising an aminoacyl tRNA analogue, whereby the protein produced by one translation system differs from the protein produced by another system at the preselected amino acid position.

21. A method of Claim 20, wherein the difference between proteins produced by the different translation systems is predetermined by the preselection of unnatural amino acid analogues attached to the aminoacyl tRNAs.

22. A method of Claim 20, wherein the protein has a molecular weight greater than about ten thousand daltons.

23. A method of Claim 20, wherein one amino acid position is substituted.

24. A method of Claim 20, wherein the unnatural amino acid substitution is selected from the group of:

i) D-phenylalanine;
ii) (S)-p-nitrophenylalanine;
iii) (S)-homophenylalanine;
iv) (S)-p-fluorophenylalanine;
v) (S)-3-amino-2-benzylpropionic acid; and
vi) (S)-2-hydroxy-3-phenylpropionic acid.

25. A method of Claim 20, wherein the mistranslation codon is a translation termination codon.

26. A method of Claim 25, wherein the translation termination codon is a UAG (amber) termination codon.

27. A method for producing an aminoacyl tRNA analogue molecule, such method comprising the steps of:

a) attaching a predetermined unnatural amino acid analogue by an aminoacyl linkage at 2' or 3' ribosyl hydroxyl positions on the 3' terminal nucleotide of a multi-nucleotide molecule (MNM); and

b) ligating the aminoacyl-multi-nucleotide molecule (aminoacyl-MNM) to a truncated tRNA molecule (tRNA(-Z)), wherein a functional aminoacyl tRNA analogue molecule is formed.

28. A method of Claim 27, wherein the multi-nucleotide molecule (MNM) corresponds to a tRNA 3' terminus.

29. A method of Claim 27 or 28, wherein the multi-nucleotide molecule (MNM) is a dinucleotide.

30. A method of Claim 29, wherein the dinucleotide is 5'-pCpA-3'.

31. A method of Claim 27, wherein ligation of the multi-nucleotide molecule (MNM) to the tRNA(-Z) molecule generates a complete tRNA molecule.

32. A method of Claim 27, wherein the tRNA(-Z) is derived from a run-off transcript.

33. A method of Claim 27 or 30, wherein the attaching of the predetermined unnatural amino acid analogue by an aminoacyl linkage at 2 or 3' ribosyl hydroxyl positions on the 3' terminal nucleotide of a multi-nucleotide molecule (MNM) is accomplished by the steps of:

a) protecting reactive chemical groups of the MNM with protective agents;

b) protecting reactive non-aminoacyl reactive groups of the amino acid analogue with a blocking agent;

c) acylating the MNM with a blocking agent-protected amino acid analogue; and

d) removing the protective agents and blocking agents from the protected reactive sites.

34. A method of Claim 33, wherein some or all of the reactive group protecting steps are substituted with steps using blocking or protective agents selected from the group consisting of:

a) o-nitrophenylsulfenyl (NSP);
b) β-cyanoethyl (EtCNO);
c) benzyloxycarbonyl (CBZ);
d) 9-fluorenylamethyloxycarbonyl (FMOC);
e) 2-(4-biphenyl) isopropyloxycarbonyl (BPOC);
f) vinyloxycarbonyl (VOC);
g) tetrahydropyranyl (THP);
h) methoxytetrahydropyranyl; and
i) photolabile groups.

35. A method of Claim 33, wherein the protecting steps are performed using o-nitrophenylsulfenyl (NPS) for both the blocking agents and protective agents.

36. A method of Claim 27, wherein the ligating of the aminoacyl-MNM to the tRNA(-Z) is performed by the enzyme T4 RNA ligase.

37. An aminoacyl tRNA analogue molecule having;

a) the formula X - A - Y - M, wherein:
X = 5' nucleotide sequence of a tRNA molecule;
A = anticodon nucleotides;
Y = 3' nucleotide sequence of a tRNA molecule;
M = amino acid analogue selected from the group consisting of:

i) modified uncharged natural amino acids;

ii) modified acidic natural amino acids; and

iii) non-alpha amino acids; and

b) activity to direct the polymerization of the M component into a nascent polypeptide chain and which can serve as an acceptor for further peptide polymerization.

38. A molecule of Claim 37, wherein the Y component has a 3' terminus of 5'-pCpCpA-3'.

39. A molecule of Claim 38, corresponding to $tRNA^{Phe}_{CUA}$ aminoacylated with (S)-p-nitrophenylalanine, wherein:

a) X comprises the 5' segment of $tRNA^{Phe}_{CUA}$ containing a "D loop" and part of an "anticodon loop";

b) A (anticodon) comprises the trinucleotide 5'-pCpUpA-3';

c) Y comprises the 3' segment of $tRNA^{Phe}_{CUA}$ containing part of an "anticodon loop", a "variable loop", a "TΨC loop", and an "acceptor stem"; and

d) M is (S)-p-nitrophenylalanine.

40. A translation system comprising an aminoacyl tRNA analogue molecule of Claim 37.

41. A coupled transcription and translation system wherein products of the transcription system are translated by a translation system comprising an aminoacyl tRNA analogue of Claim 37.

42. A substantially homogeneous protein of greater than about 10,000 daltons, wherein an unnatural amino acid has been stoichiometrically substituted at specific sites.

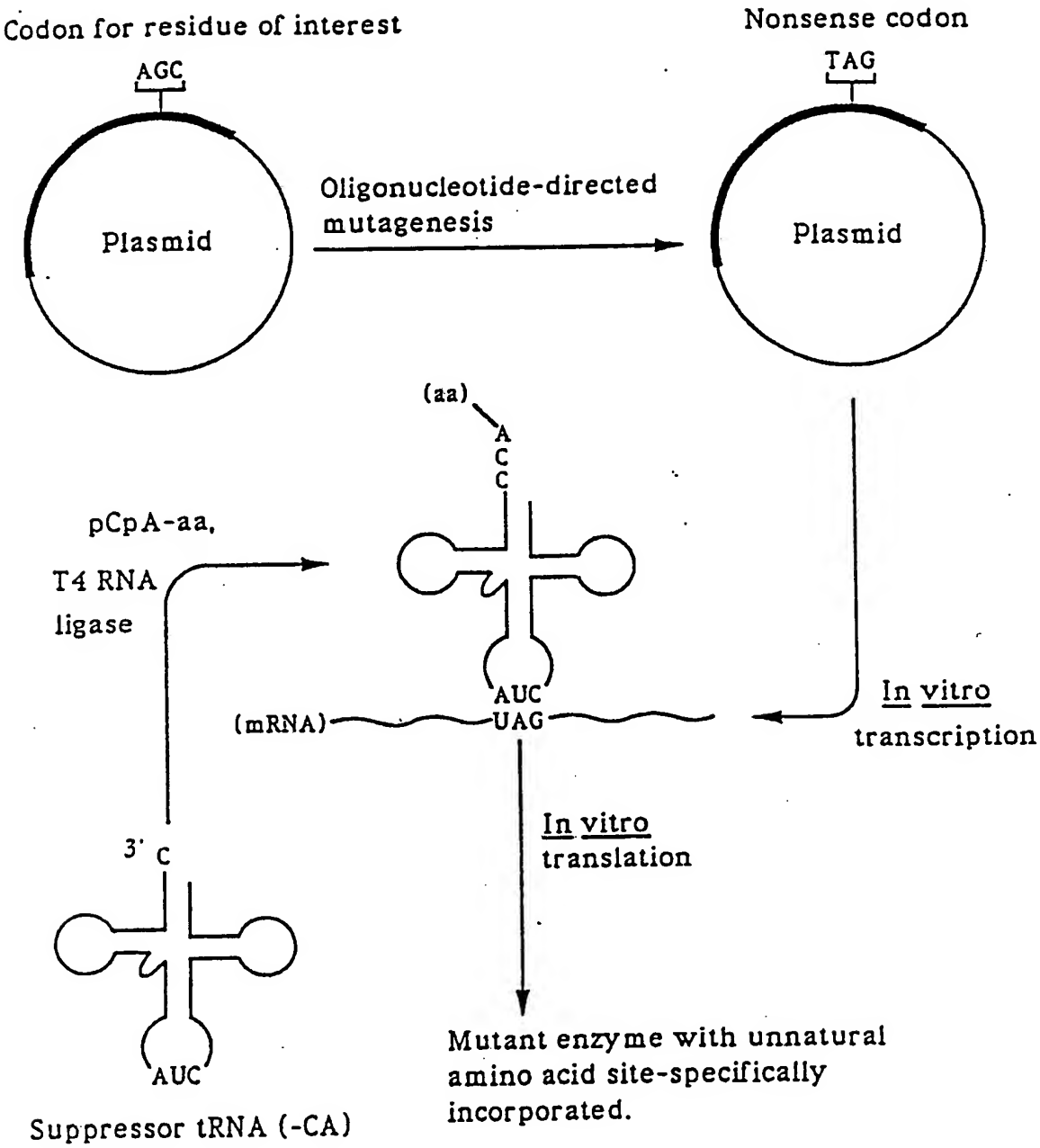

Codon for residue of interest
AGC
Plasmid
Oligonucleotide-directed mutagenesis
Nonsense codon
TAG
Plasmid
(aa)
A
C
C
pCpA-aa,
T4 RNA ligase
AUC
(mRNA)
UAG
In vitro transcription
In vitro translation
3' C
AUC
Suppressor tRNA (-CA)
Mutant enzyme with unnatural amino acid site-specifically incorporated.

FIGURE 1

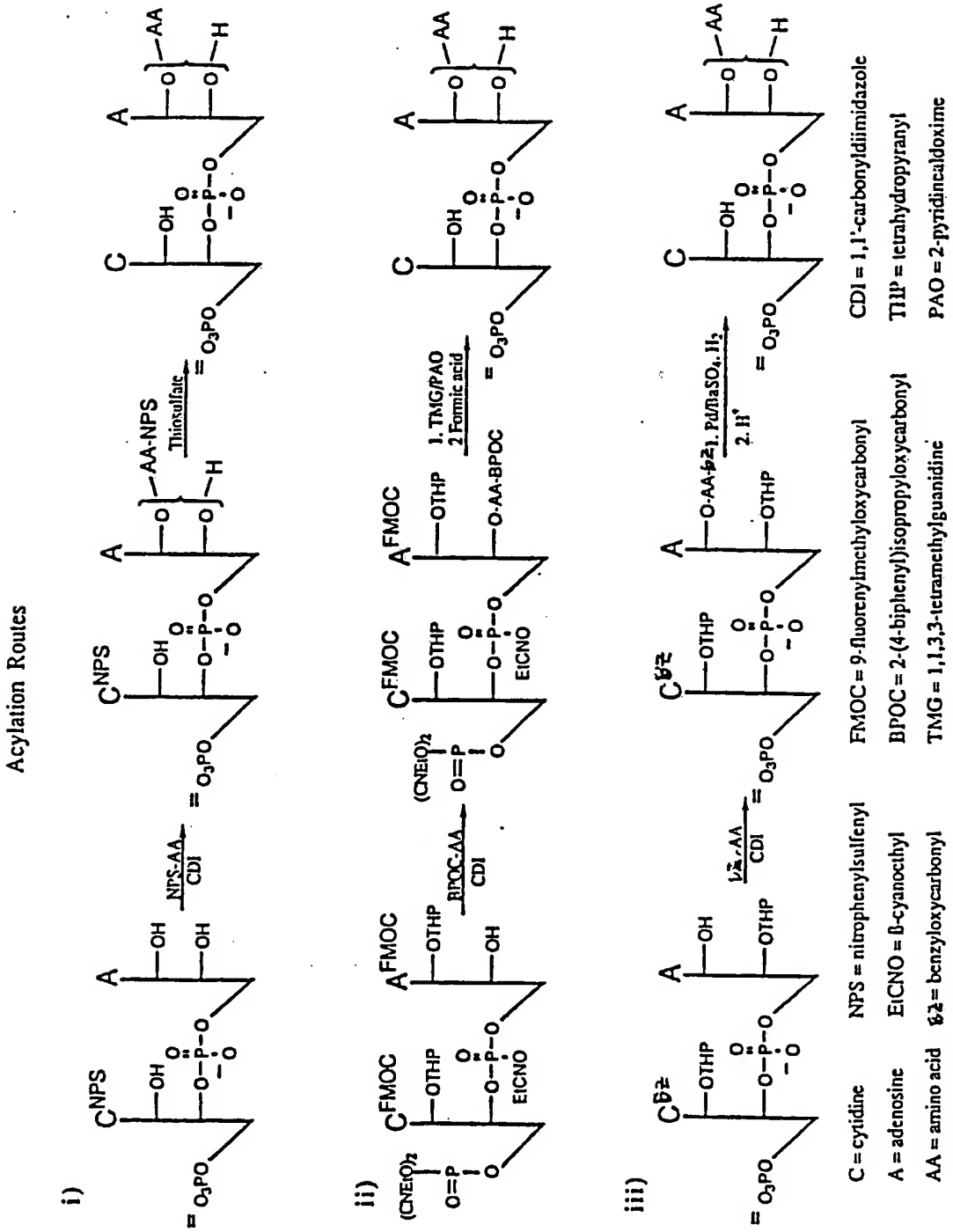

Acylation Routes
i)
ii)
iii)
NPS:AA
CDI
AA-NPS
Thiosulfate
BPOC:AA
1. TMG/PAO
2 Formic acid
Pd/BaSO4, H2
C = cytidine
A = adenosine
AA = amino acid
NPS = nitrophenylsulfenyl
EtCNO = β-cyanoethyl
Bz = benzyloxycarbonyl
FMOC = 9-fluorenylmethyloxycarbonyl
BPOC = 2-(4-biphenyl)isopropyloxycarbonyl
TMG = 1,1,3,3-tetramethylguanidine
CDI = 1,1'-carbonyldiimidazole
THP = tetrahydropyranyl
PAO = 2-pyridincaldoxime

FIGURE 2

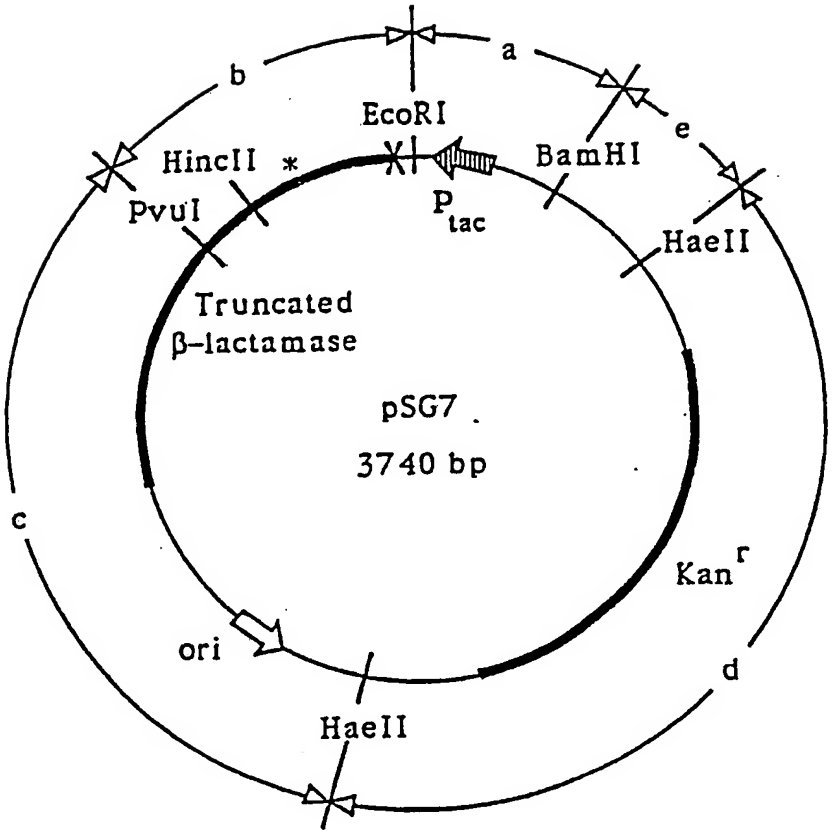

a
b
e
c
d
EcoRI
HincII
*
PvuI
BamHI
P tac
HaeII
Truncated
β-lactamase
pSG7
3740 bp
Kan r
ori
HaeII

FIGURE 3

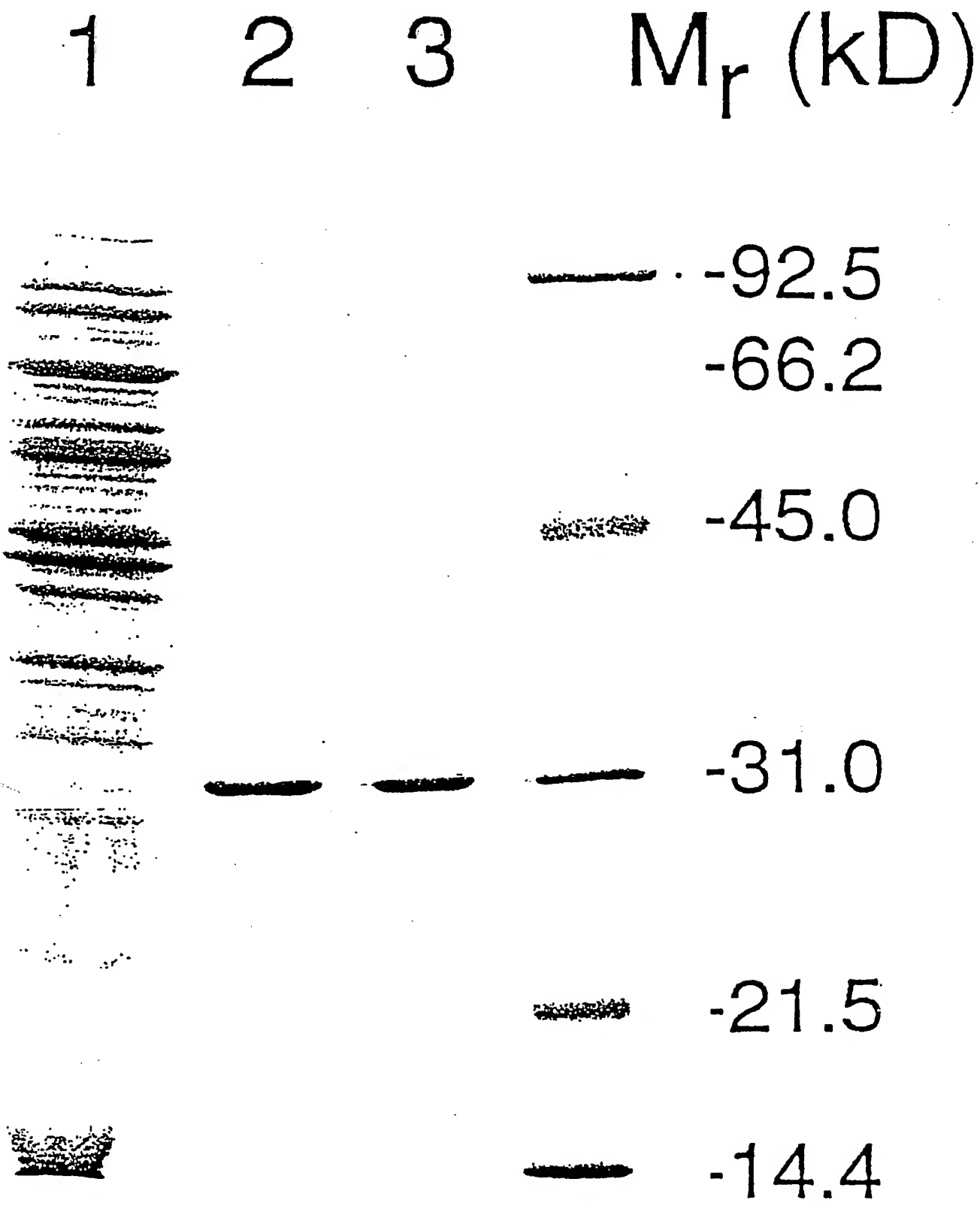


FIGURE 4

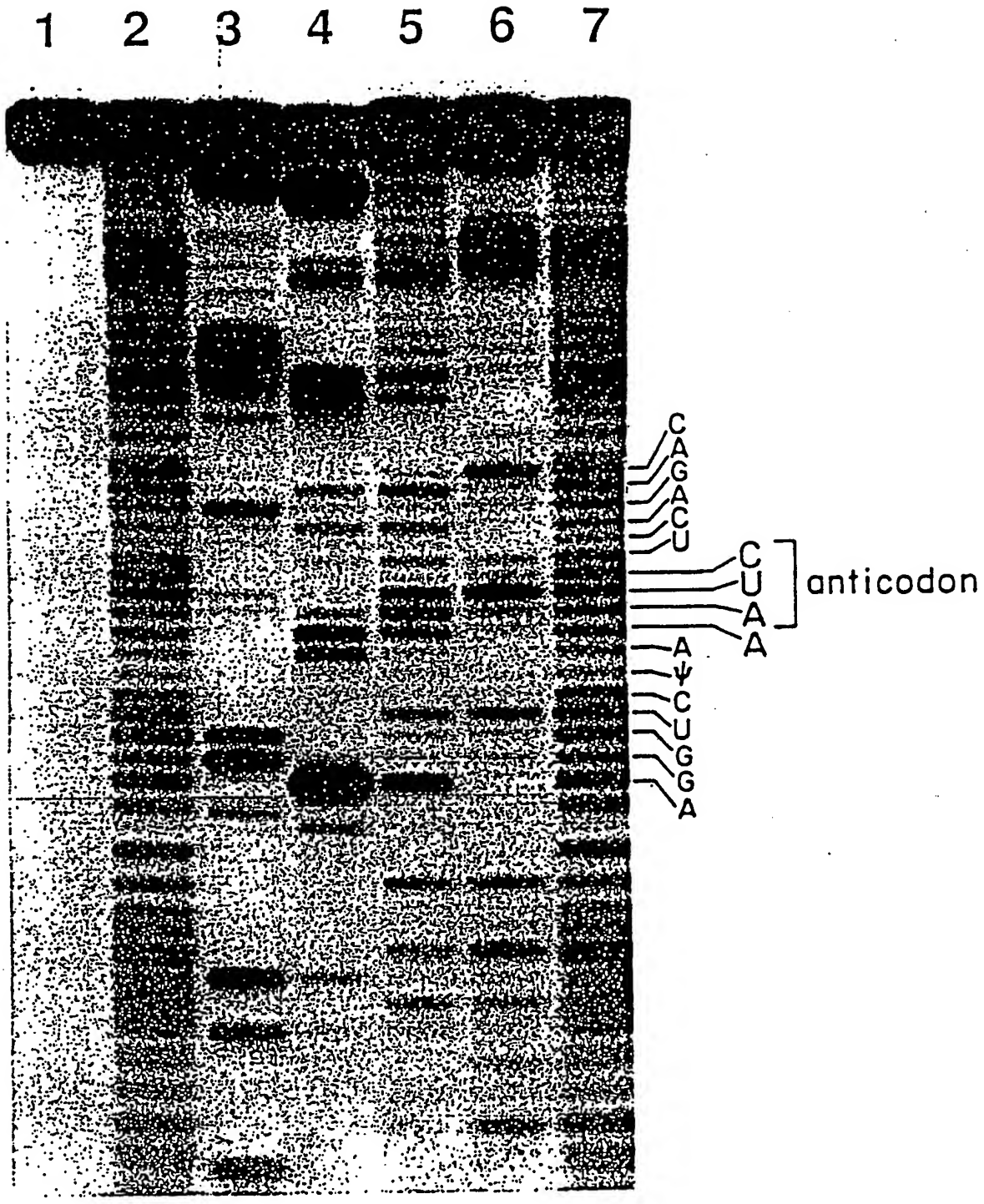


FIGURE 5

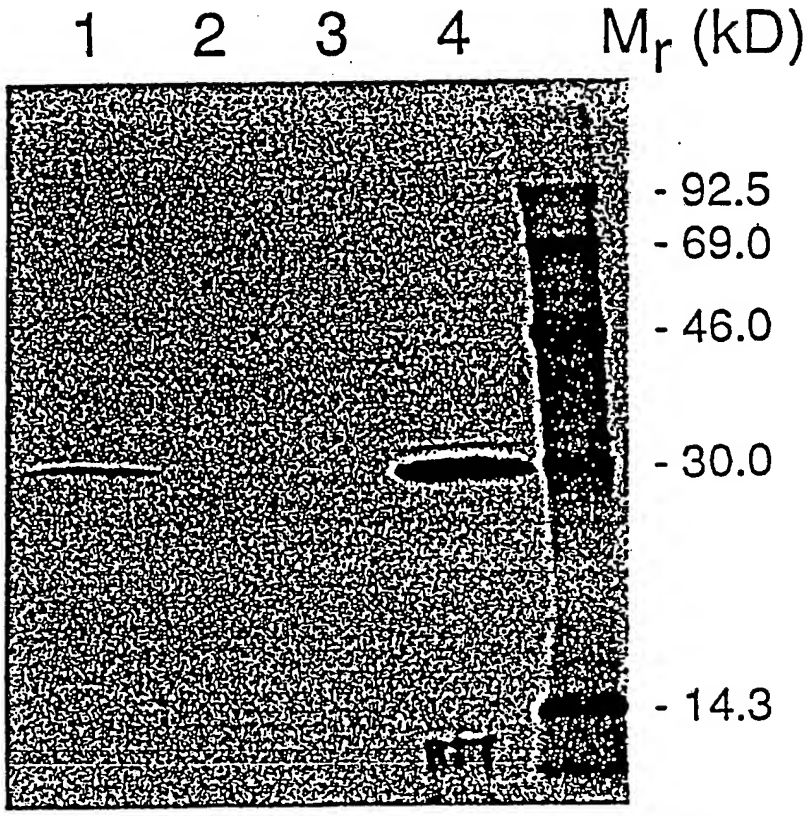


FIGURE 6

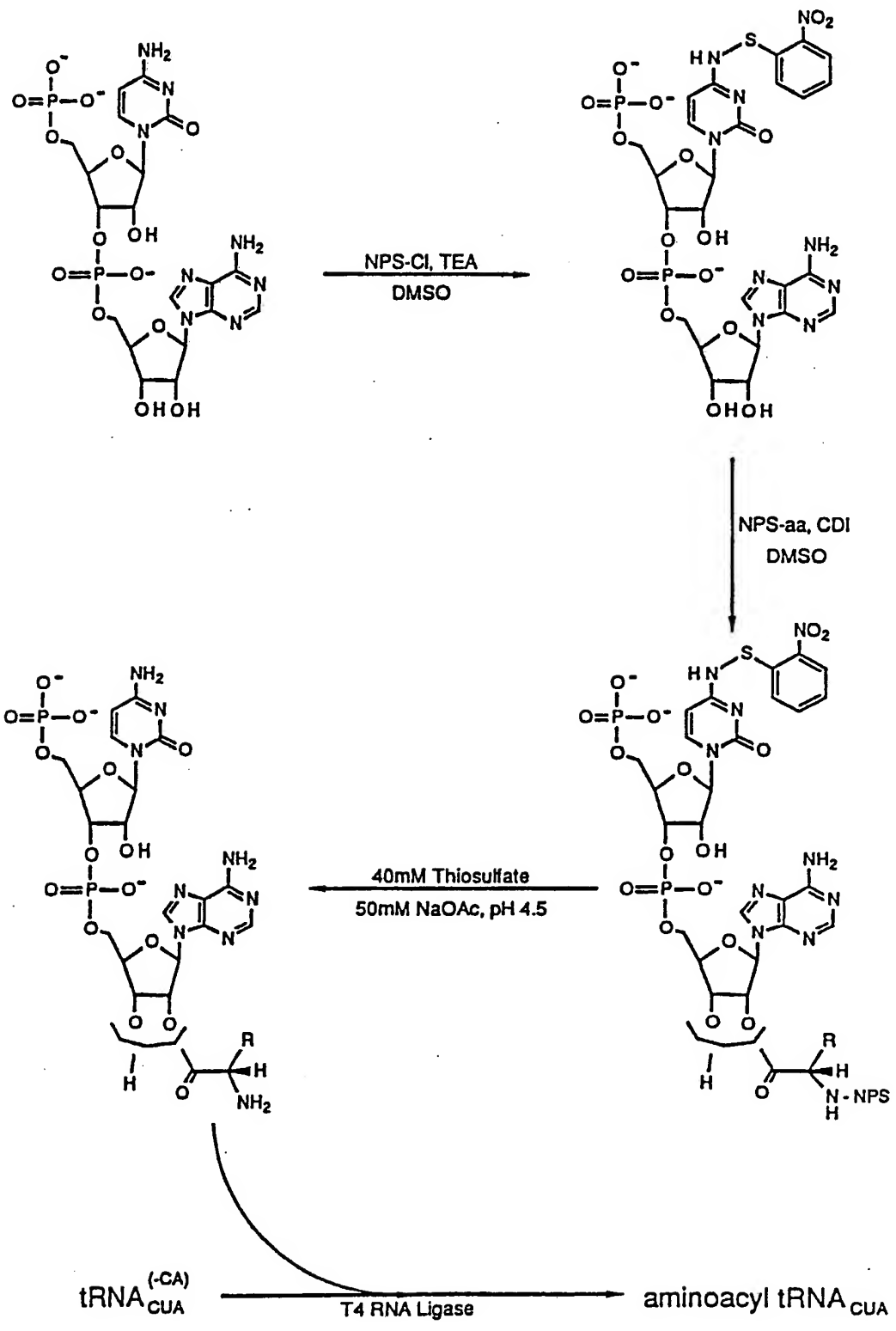

NPS-Cl, TEA
DMSO
NPS-aa, CDI
DMSO
40mM Thiosulfate
50mM NaOAc, pH 4.5
tRNA (-CA) CUA
T4 RNA Ligase
aminoacyl tRNA CUA

FIGURE 7

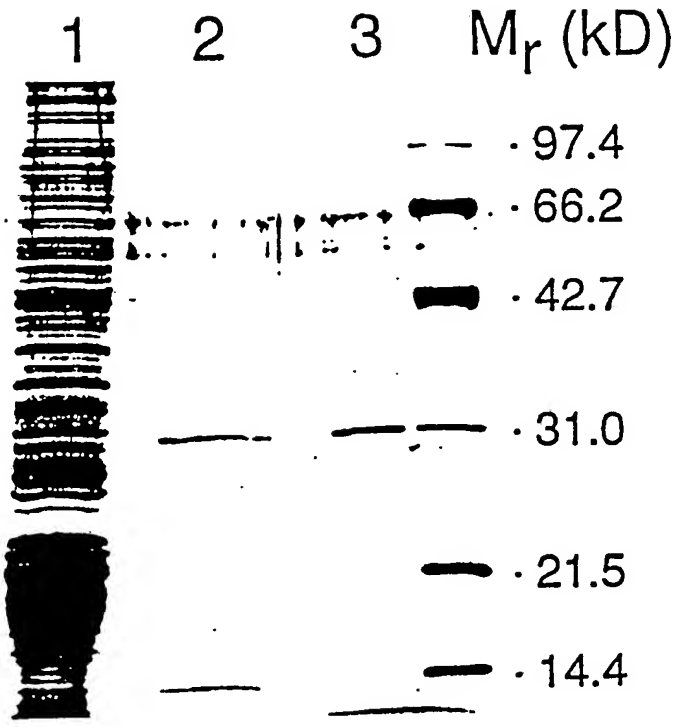
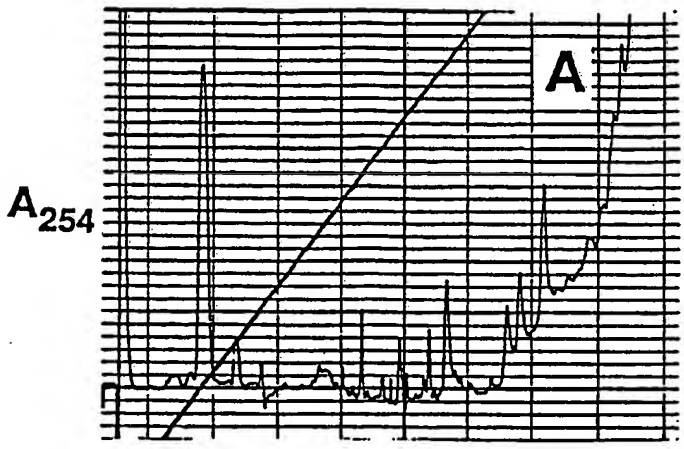
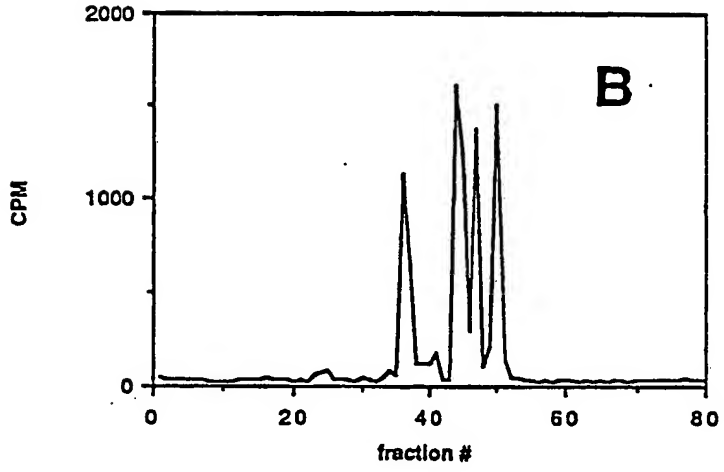


FIGURE 8



Wild type, uniformly labelled



Suppressed, labelled at Phe66

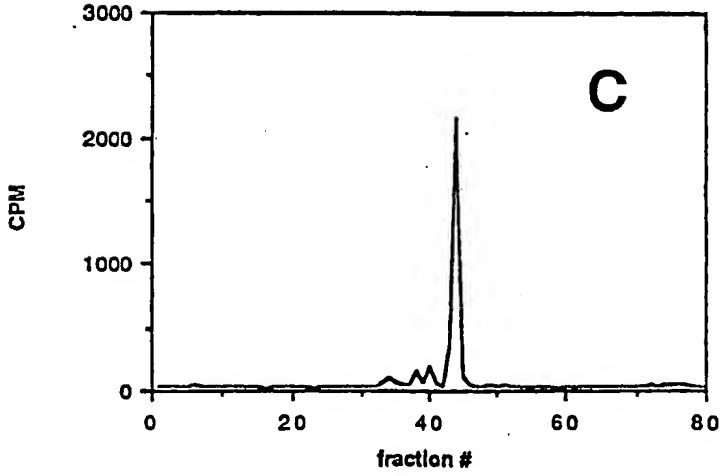


FIGURE 9

FIGURE 10

Amino acid at position 66	Plasmid	Suppressor	Enzyme Synthesized (μg/mL)	K_M (μM)	k_{cat} (s^{-1})
Phe	pSG7	-	26.0 ± 3.8†	55 ± 5	880 ± 10*
Phe	pF66am	Phe-tRNA$_{CUA}$	2.9 ± 0.9†	59 ± 6	870*
Tyr	pF66Y	-	16.9 ± 2.3†	49 ± 3	420 ± 40†
p-FPhe	pF66am	p-FPhe-tRNA$_{CUA}$	2.1 ± 0.9†	59 ± 2	1120 ± 290†
p-NO_2Phe	pF66am	p-NO_2Phe-tRNA$_{CUA}$	3.0 ± 1.0†	57 ± 4	370 ± 70†
HPhe	pF66am	HPhe-tRNA$_{CUA}$	1.0 ± 0.4†	72 ± 14	150 ± 60†
PLA	pF66am	PLA-tRNA$_{CUA}$	0†	-	-
ABPA	pF66am	ABPA-tRNA$_{CUA}$	0†	-	-
D-Phe	pF66am	D-Phe-tRNA$_{CUA}$	0†	-	-

* Based on Bradford assay quantitation of purified enzyme
† Based on incorporated radioactivity (see legend)

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/05256

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C12P 19/34 21/02; C12N 15/1; C07H 21/02

U.S. CL.: 435/69.1, 91; 530/351; 536/27

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/68.1,69.1,172.1,172.3,183+; 530/350+; 536/27+

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

APS (1975-1990) AND BIOSIS DATA BASE (1967-1990). KEYWORD: tRNA, ANALOG, AMINOACYL, AMINOACYLATION, MISACYLATION, MISINCORPORATION, CHARGING, MISCHARGING, TRANSFERASE, ANALOG, PEPTIDYLTRANSFERASE.

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
A	Annual Review Of Biochemistry, Volume 54, issued 1985(G. K. ACKERS ET AL.) "Effects Of Site Specific Amino Acid Modification On Protein Interactions And Biological Function" See pages 597-629.	1-42
A	Annual Review Of Biochemistry, Volume 48, issued 1979 (P.R. SCHIMMEL), "Aminoacyl-tRNA Synthetases: General Features And Recognition Of Transfer RNAS" See pages 601-648, particularly pages 636-640.	1-42
A	Annual Review of Genetics, Volume 19, issued 1985 (E.G. MURGOLA), "tRNA, Suppression, And The Code" See pages 57-80, particularly pages 63-65 and 69-71.	1-42

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

20 February 1990

Date of Mailing of this International Search Report

13 MAR 1990

International Searching Authority

ISA/US

Signature of Authorized Officer

LARRY MILLSTEIN

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Journal of Biological Chemistry, Volume 258, no. 7, issued 1983, April (T.G. HECKLER ET AL.) "Dipeptide Formation With Misacylated tRNAs*" see pages 4492-4495, particularly the abstract on page 4492 and the last paragraph on page 4495.	1-42
Y	Biochemistry, Volume 25, issued 1986 (J.P. ROESSER ET AL.) "Ribosome-Catalyzed Formation of an Abnormal Peptide Analogue" see pages 6361-6365, particularly pages 6361 and 6365, last paragraph.	1-42
Y	Biochemistry, Volume 27, issued 1988 (G. BALDINI ET AL.) "Mischarging Escherichia Coli tRNA With L-4'-[3-(Trifluoromethyl)-3H-Diazirin-3-Yl] Phenylalanine, a Photoactivatable Analogue of Phenylalanine", see pages 7951-7959, particularly pages 7951, 7957 and 7958.	27-39
Y	Tetrahedron, Volume 40, issued 1984 (T.G. HECKLER ET AL.) "Preparation of 2'(3')-O-Acyl-pCpA Derivatives As Substrates For T4 RNA Ligase-Mediated "Chemical Aminoacylation"", see pages 87-94, particularly page 87.	1-42
Y	Acta Chemica Scandinavica B, Volume 37, issued 1983 (J. HEIKKILA ET AL.) "The 2-Nitrophenylsulfenyl (Nps) Group For The Protection of Amino Functions of Cytidine, Adenosine, Guanosine and Their 2'-Deoxysugar Derivatives", see pages 857-864, particularly page 857.	27-39
Y	Canadian Journal of Biochemistry, Volume 58, issued 1980 (KWOK ET AL.) "Evolutionary Relationship Between Halobacterium Cutirubrum and Eukaryotes Determined by use of Aminoacyl-tRNA Synthetases as Phylogenetic Probes", see pages 213-218, particularly pages 213 and 216-218.	1-26, 40-42
Y	Proceeding Of The National Academy Of Sciences, U.S.A., Volume 79, issued 1982 (A.G. BRUCE ET AL.) "Replacement Of Anticodon Loop Nucleotides to Produce Functional tRNAs: Amber Suppressors derived From Yeast tRNA", see pages 7127-7131.	1-26, 40-42